Évaluation des interventions de santé mondiale : méthodes avancées

Valéry Ridde, Christian Dagenais

▶ **To cite this version:**

Valéry Ridde, Christian Dagenais. Évaluation des interventions de santé mondiale : méthodes avancées. Éditions Science et Bien Commun; IRD, 465 p., 2019, 978-2-7099-2766-6. 10.4000/books.irdeditions.33393 . ird-02303154

HAL Id: ird-02303154
https://ird.hal.science/ird-02303154

Submitted on 2 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Évaluation des interventions de santé mondiale

Méthodes avancées

Sous la direction de
Valéry Ridde et Christian Dagenais

esbc
éditions science et bien commun

IRD Éditions

Évaluation des interventions de santé mondiale de Valéry Ridde et Christian Dagenaisest sous une licence License Creative Commons Attribution - Partage dans les mêmes conditions 4.0 International, sauf indication contraire.

Titre : Évaluation des interventions de santé mondiale. Méthodes avancées

Sous la direction de Valéry Ridde et Christian Dagenais

Design de la couverture : Kate McDonnell, photographie de Christian Dagenais

Révision : Florence Piron, Caroline Dufresne, Élisabeth Arsenault, Catherine Chabot. Nous tenons à remercier Maria José Arauz pour sa relecture des résumés en espagnol et Segun Afolabi pour la relecture des résumés en anglais.

ISBN ePub : 978-2-924661-60-4

ISBN pour l'impression au Canada : 978-2-924661-58-1

ISBN pour l'impression en France : 978-2-7099-2766-6

Dépôt légal – Bibliothèque et Archives nationales du Québec 2019

Dépôt légal – Bibliothèque et Archive nationale Canada

Dépôt légal : 2e trimestre 2019

Ce livre est sous licence Creative Commons CC-BY-SA 4.0

Éditions science et bien commun
http://editionscienceetbiencommun.org
1085 avenue de Bourlamaque Québec (Québec) G1R 2P4 Canada

IRD Éditions, 44 bd de Dunkerque, 13572 Marseille cedex 02 France

Diffusion: info@editionscienceetbiencommun.org ou diffusion@ird.fr

Table des matières

Introduction ix
Valéry Ridde et Christian Dagenais

Partie I. La phase pré-évaluative et la pérennité

1. L'étude d'évaluabilité 3
Une intervention de prévention de l'usage de drogues à l'école au Québec
Biessé Diakaridja Soura, Jean-Sébastien Fallu, Robert Bastien et Frédéric N. Brière
2. L'évaluation de la pérennité 33
Une intervention de financement basé sur les résultats au Mali
Mathieu Seppey et Valéry Ridde

Partie II. Les approches qualitatives et participatives

3. L'évaluation qualitative, informatisée, participative et inter-organisationnelle (EQUIPO) 57
Exemple d'un programme en faveur des femmes victimes de violences en Bolivie
Mathieu Bujold et Jean-Alexandre Fortin
4. La méthode photovoix 83
Une intervention auprès de populations marginalisées sur l'accès à l'eau potable, l'hygiène et l'assainissement au Mexique
Lynda Rey, Wilfried Affodégon, Isabelle Viens, Hind Fathallah et Maria José Arauz

5. L'analyse d'une recherche-action 125
Combinaison d'approches dans le domaine de la santé au Burkina Faso
Aka Bony Roger Sylvestre, Valéry Ridde et Ludovic Queuille

Partie III. Les méthodes mixtes

6. Les revues systématiques mixtes 157
Un exemple à propos du financement basé sur les résultats
Quan Nha Hong, Anne-Marie Turcotte-Tremblay et Pierre Pluye
7. L'intégration en méthodes mixtes 187
Cadre conceptuel pour l'intégration des phases, résultats et données qualitatifs et quantitatifs
Pierre Pluye
8. La pratique de l'intégration en méthodes mixtes 213
Les multiples combinaisons des stratégies d'intégration
Pierre Pluye, Enrique García Bengoechea, David Li Tang, Vera Granikov

Partie IV. L'évaluation de l'efficacité et de l'efficience

9. Les méthodes quasi-expérimentales 241
L'effet de l'âge légal minimum sur la consommation d'alcool chez les jeunes aux États-Unis
Tarik Benmarhnia et Daniel Fuller
10. Les essais randomisés en grappe 265
Un exemple en santé maternelle et infantile
Alexandre Dumont
11. La mesure de l'équité 283
Un exemple d'intervention de gratuité des soins obstétricaux
Tarik Benmarhnia et Britt McKinnon

12. L'analyse coût-efficacité 311
Une intervention de décentralisation des soins VIH/SIDA à Shiselweni, Swaziland
Guillaume Jouquet
13. L'analyse spatiale 333
Un cas d'intervention communautaire de lutte contre le moustique Aedes aegypti au Burkina Faso
Emmanuel Bonnet, Samiratou Ouédraogo et Diane Saré

Partie V. L'évaluation des processus et de la fidélité d'implantation

14. L'analyse des processus de mise en œuvre 355
Une intervention complexe au Burkina Faso : le financement basé sur les résultats
Valéry Ridde et Anne-Marie Turcotte-Tremblay
15. L'évaluation de la fidélité d'implantation 383
Un projet de distribution d'omble chevalier aux femmes enceintes du Nunavik
Lara Gautier, Catherine M. Pirkle, Christopher Furgal et Michel Lucas
16. L'évaluation de la fidélité et de l'adaptation 403
Un exemple de mise en œuvre des interventions en santé mondiale
Dennis Pérez, Marta Castro et Pierre Lefèvre
17. L'évaluation réaliste 433
L'exemple de l'adoption d'une politique publique de santé au Bénin
Jean-Paul Dossou et Bruno Marchal

À propos de la maison d'édition 463

Introduction

VALÉRY RIDDE ET CHRISTIAN DAGENAIS

Il y a dix ans, nous étions interpellés par le fait que nous faisions souvent face, lors de nos enseignements de l'évaluation de programmes à des francophones, au défi de disposer d'écrits pour soutenir nos échanges pédagogiques. À l'époque, les bibliothèques regorgeaient d'ouvrages de référence en anglais dont nous nous servions très souvent pour leurs vertus pédagogiques (Patton 1997; Weiss 1998; Chen 2005), mais nous butions toujours sur les défis linguistiques lorsque nous enseignions à des francophones. C'est pourquoi nous nous sommes lancés dans une première aventure éditoriale avec la publication en 2009 d'un ouvrage pédagogique en français pour former des étudiant-e-s et des professionnel-le-s[1] à l'évaluation de programmes. Au-delà de la langue, l'originalité de l'ouvrage était son approche pluridisciplinaire et intersectorielle. En effet, après avoir présenté des propositions de synthèse conceptuelle (y compris dans des textes d'autrices et d'auteurs anglophones importants que nous avions traduits), le livre proposait de nombreux exemples d'évaluation d'intervention. Ces exemples se situaient dans plusieurs secteurs (éducation, santé, action sociale, etc.) et montraient de manière didactique comment plusieurs méthodes (quantitatives, qualitatives, mixtes) et disciplines pouvaient être mobilisées pour répondre à des questions d'évaluation utiles aux intervenant-e-s et décideurs et décideuses. Ce premier ouvrage a été apprécié par les lecteurs, les lectrices et les enseignant-e-s et nous avons ainsi senti le besoin, en 2012, d'en publier une édition revue et augmentée (Ridde et Dagenais 2012). En outre, il a été traduit en espagnol et diffusé en Amérique latine par la maison d'édition d'une université colombienne qui souhaitait mieux diffuser ces approches sur ce continent, confirmant le

1. Respectant les valeurs des Éditions science et bien commun, cette introduction, comme tous les chapitres du livre, suit les principes de l'écriture inclusive et propose de nouvelles graphies ou des tournures épicènes afin de lutter contre l'invisibilité des femmes ou du féminin. En particulier, il applique l'accord de proximité plutôt que la victoire du masculin et utilise les doublets ou les tirets selon la forme du féminin. Il contribue aussi aux retrouvailles de la langue française avec des formes grammaticales féminines ayant été effacées de l'usage par l'Académie française à la Renaissance.

besoin de disposer d'ouvrages pédagogiques dans des langues autres que l'anglais.

Mais au fil des années et de notre expérience de recherche et d'enseignement dans les pays du Sud, nous avons senti un triple besoin qui est à l'origine du présent ouvrage collectif.

D'une part, nous avons constaté la nécessité de continuer à publier en français face à la permanence des difficultés des étudiant-e-s et des collègues œuvrant uniquement dans un espace francophone et ne maîtrisant pas facilement l'anglais. Si quelques livres ont été publiés en français depuis 2009 dans le domaine de l'évaluation (Larivière et Corbière 2014; Hurteau, Houle, et Guillemette 2012), aucun ne disposait encore d'une spécificité propre au contexte du Sud, ou plus exactement des pays à faible revenu, pour éviter tout amalgame d'une uniformité contextuelle. En 2016, alors que nous avions déjà lancé la production du présent ouvrage, un autre livre spécifique à la pratique de l'évaluation en Afrique a été publié, mais il portait moins sur les méthodes que sur des pratiques réflexives (Ridde *et al.* 2016). Nous reviendrons sur cette spécificité contextuelle, mais souhaitons ici montrer l'importance de rendre accessibles en français des écrits qui concernent un champ encore majoritairement dominé par le monde anglophone.

D'autre part, si notre premier ouvrage constitue, pensons-nous, une très belle introduction à l'évaluation des interventions pour les novices dans ce domaine, il nous est vite apparu la nécessité d'en approfondir les contours et d'entrer dans la complexité de méthodes avancées. Or, il nous est aussi très vite apparu que de tels ouvrages interdisciplinaires en français et en santé mondiale n'existaient pas. Pourtant, notamment pour celles et ceux qui s'engagent dans des processus d'évaluation rigoureux dépassant les seules consultations répondant à des demandes formulées par un organisme client (ce qui est aussi évidemment un mandat important), il devient essentiel de maitriser des méthodes complexes permettant de répondre à des questions d'évaluation complexes ciblant des programmes tout aussi complexes.

Cette complexité rend impossible dans bien des cas de mettre en place ce que certain-e-s économistes du développement considèrent comme l'étalon de l'analyse de la causalité, soit les devis contrôlés-randomisés qui, la plupart du temps, sont insuffisants pour répondre aux questions soulevées dans le contexte du Sud. Cette complexité entraine aussi de grands défis

évaluatifs dans le cas d'interventions répondant aux caractéristiques de la complexité (Robert et Ridde 2014). Mais force est de constater le besoin de mobiliser des méthodes avancées pour de telles analyses et donc de les présenter. En outre, nous pensons que le contexte des pays à faible revenu est favorable à l'innovation méthodologique et à la mise en branle de telles méthodes.

On ne reviendra pas sur les sempiternels débats autour du concept de santé mondiale (Taylor 2018), car nous souhaitons ici surtout retenir le fait qu'il s'inscrit dans une présence de ressources moindres et de fortes inégalités, mais aussi d'enjeux de pouvoir exacerbés par la forte présence d'une aide publique au développement des pays à revenu élevé et d'une collaboration internationale dans la pratique de l'évaluation. Puisque les contextes sont omniprésents et influencent toutes les « étapes » d'une évaluation (Craig *et al.* 2018), il est essentiel d'en tenir compte pour choisir les méthodes pertinentes scientifiquement, mais contextuellement aussi. Ainsi, contrairement aux idées reçues, c'est dans un contexte de rareté des ressources que les évaluateurs et évaluatrices se doivent d'être le plus agiles, d'innover et de tenter de déployer des méthodes appropriées pour atteindre leurs objectifs évaluatifs. Les chapitres réunis dans cet ouvrage sont de parfaits exemples de cette agilité et de la manière dont il est possible de déployer des approches complexes pour évaluer des interventions. Personne n'a attendu les récentes tentatives de créer un concept nouveau autour de la recherche interventionnelle en santé publique pour évaluer de manière rigoureuse les interventions en santé mondiale. Les auteurs et les autrices que nous avons mobilisés pour bâtir cet ouvrage font la preuve de ce dynamisme méthodologique, y compris dans un contexte où les moyens mis à la disposition de l'évaluation sont largement moindres que ce qui est disponible au Nord. Mais la diffusion de ces innovations, pour paraphraser Rogers, est encore trop souvent cantonnée à quelques expert-e-s… surtout anglophones et disposant des moyens d'y avoir accès.

Ainsi, et enfin, le troisième besoin que nous avons identifié et auquel ce livre tente de répondre, est celui de la diffusion libre des savoirs, qui est un enjeu majeur en santé mondiale (Smith *et al.* 2017). À quoi peut-il bien servir de développer toutes ces innovations méthodologiques si elles restent confinées et accaparées par les personnes qui maîtrisent l'anglais et qui ont les moyens d'y accéder? Comment faire en sorte que les étudiants et étudiantes francophones basés au Sud puissent avoir accès à des

connaissances originales et utiles alors qu'ils et elles n'ont pas les moyens de s'offrir des livres, même accessibles en version payante sur Internet? Comment soutenir le développement du réseau des évaluateurs émergents francophones[2]? C'est ici que notre rencontre avec le projet des Éditions science et bien commun et son souhait d'une démocratisation des savoirs et d'une justice sociale des connaissances nous a agréablement interpellés. Nous avions déjà été en mesure de publier un de nos ouvrages sur les idées reçues en santé mondiale en version gratuite grâce à une initiative des Presses de l'Université de Montréal (Ridde et Ouattara 2015). Nous avions aussi subventionné une reproduction de la première édition de notre livre d'introduction à l'évaluation pour le publier en Afrique de l'Ouest à un tarif réduit, en collaboration avec les Éditions L'Harmattan Burkina Faso. Mais ces subventions sont temporaires et nous voulions vraiment que notre nouvel ouvrage soit le plus largement accessible. La communauté des personnes qui réalisent des évaluations dans le monde francophone se développe à l'échelle mondiale, tous les pays d'Afrique francophone offrent au moins une maitrise de santé publique où l'évaluation commence à être enseignée… voici autant d'opportunités que cet accès libre à des méthodes avancées rendra utile, croyons-nous.

Ainsi, ce livre s'adresse à tous les amoureux de l'évaluation (oui, oui, cela existe!) qui souhaitent développer leurs connaissances et compétences en méthodes avancées. Il s'adresse d'abord aux personnes qui agissent dans un contexte de santé mondiale, mais pas uniquement. En effet, si les cas présentés concernent le domaine de la santé, au sens large du terme et notamment au sens populationnel, il est évident que les méthodes que nous présentons peuvent être déployées dans d'autres domaines (éducation, économie, etc.) et dans d'autres contextes que ceux des pays du Sud. Chacun-e y trouvera donc son compte et saura user de son agilité pour adapter et transposer ces méthodes ailleurs, dans d'autres champs et d'autres circonstances.

2. http://www.portail-rfe.org/RF-Ee/qui-somme-nous

Le plan du livre

L'ouvrage est composé de cinq sections qui représentent autant de grandes dimensions méthodologiques.

La première section est très innovante, car elle présente des méthodes rarement abordées dans les ouvrages et articles francophones en évaluation. Notre approche générale s'inscrit dans une démarche d'évaluation fondée sur la théorie des interventions (Chen 2005; Weiss 1998). Dans ce sens, il devient essentiel de maîtriser ce que Beaudry et Gauthier appellent la phase pré-évaluative (Beaudry et Gauthier, 1992), autrement dit l'analyse de « l'évaluabilité » des interventions. Les différentes étapes du processus d'une étude « d'évaluabilité » sont présentées dans le chapitre 1 (Soura et collègues). Le cas présenté porte sur un programme de prévention en milieu scolaire et vise à déterminer jusqu'à quel point ce programme est évaluable.

La pérennité des interventions est aussi un thème relativement délaissé par les ouvrages francophones. Lorsqu'elle est abordée, elle devient souvent la dernière roue du carrosse. Pourtant, les recherches nous disent depuis longtemps que la pérennité devrait devenir la première préoccupation des intervenant-e-s, et non la dernière, lorsque l'on planifie une intervention (Pluye *et al.* 2005). Nous avons donc décidé de l'intégrer dans la première partie, car le premier auteur de ce chapitre portant sur une intervention au Mali se plaignait d'être relégué, comme la pérennité, à la fin de l'ouvrage. Il a eu raison et nous le remercions de son plaidoyer. La pérennisation des interventions en Afrique est trop peu étudiée. Ce chapitre présente une méthodologie pour y arriver (Seppey et Ridde, chapitre 2).

La deuxième section comporte trois chapitres qui présentent des approches d'évaluation usant de méthodes qualitatives. Si cet usage n'a rien d'innovant dans le domaine de la santé mondiale (Belaid *et al.* 2016), nous voulions montrer quelques approches originales et peu diffusées en français. Ces trois chapitres proposent donc des processus intéressants où le qualitatif est mobilisé de manière rigoureuse. Le premier (Bujold et Fortin, chapitre 3) porte sur une l'évaluation participative d'un programme visant l'égalité hommes-femmes en Bolivie . Elle met à contribution une méthode systématique en trois volets (qualitatif, participatif, inter-organisationnel). Le deuxième chapitre de cette section (Rey et collègues, chapitre 4) décrit une évaluation menée au Mexique qui mobilise une autre approche

qualitative participative utilisant la photo comme support d'expression : le photovoix. Elle porte sur les perceptions de jeunes autochtones de leur participation à un programme basé sur les arts pour favoriser le changement de comportements. Enfin, le troisième chapitre de cette section (Aka et collègues, chapitre 5) porte sur l'évaluation d'une démarche de recherche-action au Burkina Faso. L'approche d'évaluation adoptée combine une analyse externe à une analyse réflexive afin de produire des résultats permettant de guider d'autres recherches-action.

La troisième section offre l'opportunité de réaliser notre souhait de diffuser plus largement le recours aux méthodes mixtes pour l'évaluation (Bujold *et al.* 2018). Trois chapitres permettent de montrer comment réaliser des revues mixtes, ainsi que l'intégration des données qui est au cœur de la pertinence de ces méthodes. Des exemples utiles en santé mondiale facilitent cette démonstration. Le premier de ces chapitres (Hong et collègues, chapitre 6) porte sur les revues mixtes qui combinent des études quantitatives, qualitatives et mixtes en utilisant une approche systématique. Il présente un exemple de revue mixte tiré d'un projet de financement basé sur les résultats au Burkina Faso; quatre devis de synthèse sont présentés. Le deuxième chapitre de cette section (Pluye, chapitre 7) propose un cadre conceptuel incluant tous les types d'intégration des méthodes qualitatives et quantitatives et les diverses combinaisons possibles de ces stratégies. Le chapitre suivant (Pluye et collègues, chapitre 8) illustre quant à lui plusieurs de ces types de combinaisons à l'aide d'études en santé mondiale.

La quatrième section est consacrée à l'évaluation de l'efficacité et de l'efficience des interventions. Les trois premiers chapitres de cette section explicitent les démarches méthodologiques pour vérifier à l'aide de méthodes statistiques complexes que les objectifs des interventions ont été atteints. Le premier (Benmarhnia et Fuller, chapitre 9) présente deux méthodes quasi-expérimentales applicables lorsque les essais contrôlés randomisés sont impossibles à organiser : la méthode de Différence dans les Différences (DD) et l'approche de Régression avec Discontinuité (RD). L'application de ces deux méthodes est utilisée pour mesurer les effets de l'âge légal minimum sur la consommation d'alcool des jeunes. Le deuxième chapitre de cette section (Benmarhnia et McKinnon, chapitre 10) présente différentes méthodes de mesure des inégalités de santé d'une intervention d'exemption de paiement des frais de soins. Ces mesures incluent notamment l'indice de concentration et l'indice de Theil. Le troisième

chapitre de la section (Dumont, chapitre 11) présente les modèles statistiques appropriés et les défis méthodologiques liés aux essais randomisés en grappe utilisés pour évaluer l'impact d'une intervention visant à réduire la mortalité maternelle et néonatale hospitalière au Mali et au Sénégal. Le quatrième chapitre de cette section (Jouquet, chapitre 12) présente une analyse coût-efficacité d'une intervention VIH réalisée au Swaziland.

Enfin, la cinquième section présente plusieurs démarches portant toutes sur l'analyse des processus et de la fidélité d'implantation dans des contextes variés, du Nunavik à Cuba en passant par l'Afrique de l'Ouest et jusqu'à l'Afrique australe. Sur le plan des méthodes, cette dernière section est innovante en proposant des approches rarement présentées ou utilisées pour le moment : les analyses spatiales (Bonnet, chapitre 13), une étude de cas multiples utilisant un processus pyramidal d'analyse (Ridde et Turcotte-Tremblay, chapitre 14), l'évaluation de la fidélité (Pérez et collègues, chapitre 15 ; Gauthier et collègues, chapitre 16) ou encore l'approche réaliste (Dossou et Marchal, chapitre 17).

Nous ne pouvions terminer cette introduction sans remercier l'ensemble des protagonistes de cet ouvrage. D'abord, les multiples auteurs et autrices et leurs collaborateurs et collaboratrices qui ont patiemment répondu à nos exigences éditoriales didactiques et accepté de participer à une publication dont la durée de production a été proportionnelle à leur motivation. Ensuite, Linda François qui nous a aidés dans la mise en forme des manuscrits des chapitres et qui a assuré le suivi avec les auteurs et les autrices pour répondre aux exigences éditoriales. L'une de leurs requêtes était de disposer de résumé des chapitres dans le maximum de langues possibles. Nous devons ainsi remercier celles et ceux qui ont bien voulu nous aider dans ces traductions. Enfin, il nous faut remercier les Éditions science et bien commun pour leur désir de publier en accès libre cet ouvrage et l'ensemble de leur soutien durant sa production, ainsi que le soutien financier de l'Institut de Recherche pour le Développement (IRD) et de l'Équipe RENARD.

Références

Beaudry, J. et Gauthier, B. (1992). L'évaluation de programme. Dans B. Gauthier (dir.), *Recherche sociale : de la problématique à la collecte de données* (2e éd., p. 425-452). Québec : Presses de l'Université du Québec.

Belaid, L., Bodson, O., Ridde, V., Samb, O. et Turcotte-Tremblay (2016).

Innovations qualitatives dans l'évaluation des interventions en santé. Dans J. Kivits, F. Balard, C. Fournier et M. Winance (dir.), *Les recherches qualitatives en santé*. Malakoff : Armand Colin.

Bujold, M., Hong, Q. N., Ridde, V., Bourque, C. J., Dogba, M. J., Vedel, I. et Pluye, P. (2018). *Oser les défis des méthodes mixtes en sciences sociales et sciences de la santé*. Montréal : 117e Cahiers scientifiques de l'ACFAS.

Chen, H.-T. (2005). *Practical program evaluation : assessing and improving planning, implementation, and effectiveness*. Thousand Oaks, CA: Sage Publications.

Craig, P., Di Ruggiero, E., Frohlich, K. L., Mykhalovskiy, E. et White, M., on behalf of the Canadian Institutes of Health Research (CIHR)–National Institute for Health Research (NIHR) Context Guidance Authors Group (listed alphabetically), R. Campbell et al. (2018). *Taking account of context in population health intervention research: guidance for producers, users and funders of research*.
https://doi.org/10.3310/CIHR-NIHR-01

Hurteau, M., Houle, S. et Guillemette, F. (dir.). (2012). *L'évaluation de programme axée sur le jugement crédible*. Québec : Presses de l'Université du Québec.

Larivière, N., et Corbière, M. (dir.). (2014). *Méthodes qualitatives, quantitatives et mixtes : dans la recherche en sciences humaines, sociales et de la santé*. Québec : Presses de l'Université du Québec.

Patton, M. Q. (1997). *Utilization-focused evaluation* (3e éd.). London: Sage Publications.

Pluye, P., Potvin, L., Denis, J.-L., Pelletier, J. et Mannoni, C. (2005). Program sustainability begins with the first events. *Evaluation and program planning*, 28, 123-137.

Ridde, V. et Dagenais, C. (dir.). (2012). *Approches et pratiques en évaluation de programme* (2e éd.). Montréal : Presses de l'Université de Montréal.

Ridde, V., Kouanda, S., Kobiané, J.-F. et Zongo, S. (dir.). (2016). *Pratiques et méthodes d'évaluation en Afrique*. Paris : L'Harmattan.

Ridde, V. et Ouattara, F. (2015). *Des idées reçues en santé mondiale*. Montréal : Presses de l'Université de Montréal.

Robert, E. et Ridde, V. (2014). L'approche réaliste pour l'évaluation de

programmes et la revue systématique : de la théorie à la pratique. *Mesure et évaluation en éducation*, 36(3), 79-108.

Smith, E., Haustein, S., Mongeon, P., Shu, F., Ridde, V. et Larivière, V. (2017). Knowledge sharing in global health research – the impact, uptake and cost of open access to scholarly literature. *Health Research Policy and Systems*, 15: 73
https://doi.org/10.1186/s12961-017-0235-3

Taylor, S. (2018). "Global Health": meaning what? *BMJ Global Health*, 3(2), e000843.

Weiss, C. H. (1998). *Evaluation : methods for studying programs and policies* (2e éd.). Upper Saddle River, NJ : Prentice Hall.

Citation

Ridde, Valery et Christian Dagenais (2019). Introduction. In *Évaluation des interventions de santé mondiale. Méthodes avancées*. Sous la direction de Valéry Ridde et Christian Dagenais, pp. ix-xvii. Québec : Éditions science et bien commun et Marseille : IRD Éditions.

PARTIE I

LA PHASE PRÉ-ÉVALUATIVE ET LA PÉRENNITÉ

1. L'étude d'évaluabilité

Une intervention de prévention de l'usage de drogues à l'école au Québec

BIESSÉ DIAKARIDJA SOURA, JEAN-SÉBASTIEN FALLU, ROBERT BASTIEN ET FRÉDÉRIC N. BRIÈRE

L'évaluation de la pérennité

Définition de la méthode

L'étude d'évaluabilité est un processus rigoureux de collecte et d'analyse de données permettant de s'assurer que l'évaluation du programme est réalisable et que celle-ci pourra produire des résultats satisfaisants et utiles aux parties prenantes.

Forces de la méthode

- Processus participatif permettant de faire contribuer les parties prenantes à l'amélioration du programme et de l'évaluation.
- Permet de mieux comprendre la théorie du programme et de développer son modèle logique
- Permet de sauver du temps et des ressources en recommandant les modalités de la conduite d'une évaluation.
- Permet une meilleure planification de l'évaluation proprement dite.

Défis de la méthode

- Peut prendre beaucoup de temps si les différentes parties prenantes ne s'entendent pas ou si les activités ne sont pas bien organisées ou encore s'il y a un grand nombre de personnes à consulter.
- Peut accroître les mésententes et provoquer des frustrations chez certaines parties prenantes.

Pour Craig et Campbell (2015), l'étude d'évaluabilité (ÉÉ) est une approche systématique destinée à planifier l'évaluation des interventions ou programmes. Une des contributions les plus significatives de l'étude d'évaluabilité réside dans le fait qu'elle peut fournir des recommandations en vue de l'amélioration du programme, de la bonne conduite de l'implantation et de l'évaluation ultérieure qui devrait s'ensuivre. Alors qu'elle a été développée depuis les années 1970 pour répondre aux difficultés et défis dans l'évaluation de programmes gouvernementaux aux États-Unis (Wholey

1976), ce n'est que récemment que l'utilisation de l'étude d'évaluabilité a connu une recrudescence tant chez les praticiens e praticiennes de l'évaluation que dans les milieux universitaires (Trevisan et Yi 2003). Selon Trevisan et Walser (2014), ce regain d'intérêt a coïncidé avec la demande croissante de reddition de compte aux gestionnaires des programmes.

L'étude d'évaluabilité est une évaluation de type exploratoire rapide pouvant se réaliser dans une période plus ou moins courte, allant de quelques semaines à plusieurs mois (Davies 2013; Wholey 2010). Elle comporte plusieurs étapes permettant de collecter des informations sur les intentions des initiateurs du programme, les objectifs poursuivis, les types de données à collecter et les modalités d'utilisation des résultats de l'évaluation. La réalisation d'une étude d'évaluabilité peut également permettre de développer le modèle logique du programme, un outil nécessaire à la conduite de l'évaluation proprement dite.

Bref, l'étude d'évaluabilité s'avère un exercice utile et pertinent pour l'évaluation de programmes et les décideurs (Soura, Dagenais, Bastien, Fallu et Janosz 2016). Cependant, elle reste peu utilisée, notamment dans le milieu francophone. D'ailleurs, on dénombre plusieurs termes utilisés en français pour évoquer cette approche.

L'objectif de ce chapitre est d'abord de présenter succinctement l'étude d'évaluabilité pour ensuite décrire un exemple de son utilisation dans le cadre de l'évaluation d'une intervention de prévention de l'usage de substances psychoactives en milieu scolaire, avant d'amorcer une analyse réflexive de notre expérience.

Présentation de l'étude d'évaluabilité

L'étude d'évaluabilité est un processus rigoureux de collecte et d'analyse de données permettant de s'assurer que l'évaluation du programme est réalisable et que celle-ci pourra produire des résultats satisfaisants et utiles aux parties prenantes (Kaufman-Levy et Poulin 2003). On entend par partie prenante toutes les personnes ou entités ayant un intérêt particulier à la mise en œuvre d'un programme et à l'évaluation de celui-ci. L'étude d'évaluabilité peut également s'entrevoir comme une évaluation exploratoire ou une activité pré-évaluative qui devrait permettre la mise en place de conditions optimales à une évaluation formative ou sommative formelle

(Leviton, Khan, Rog, Dawkins, et Cotton 2010a). Les premières ébauches de l'étude d'évaluabilité remontent aux années 1970, lorsqu'un groupe de chercheuses et chercheurs du *Urban Institute* de Washington aux États-Unis mit en place cette approche en réponse aux difficultés et défis rencontrés dans l'évaluation des programmes gouvernementaux. Selon ces chercheuses et chercheurs, lesdits programmes étaient insuffisamment mis en œuvre ou trop peu matures pour être évalués ou encore avaient des objectifs jugés irréalistes (Burrows, Bilodeau et Litvak 2012; Leviton, Khan, Rog, Dawkins et Cotton 2010 b; Nay et Kay 1982; Wholey 1976). Les critiques portaient également sur le décalage constaté entre le programme évalué et les résultats de cette évaluation (Leviton *et al.* 2010b). S'est alors engagée une réflexion sur la capacité des évaluations, selon la façon dont elles étaient conduites, à informer les preneurs et preneuses de décision dans l'élaboration des politiques publiques et la nécessité de dépenser des montants élevés pour des interventions qui auraient très peu d'effets sur les populations ciblées (Trevisan et Yi 2003; Van Voorhis et Brown 1987). C'est dans ce contexte que Horst et ses collègues entreprirent d'identifier les causes profondes de l'échec des programmes et de l'incapacité de l'évaluation à satisfaire les besoins des parties prenantes (Horst, Nay, Scanlon et Wholey 1974). Ces derniers recommandèrent alors la conduite d'un *Pre-assessment of Evaluability* qui permettrait d'accroître la capacité d'utilisation des résultats de l'évaluation. C'est cette activité pré-évaluative qui fut plus tard dénommée *evaluability assessment* ou étude d'évaluabilité ou encore évaluation de l'évaluabilité.

La conduite d'une étude d'évaluabilité consiste à suivre des étapes décrites dans les travaux réalisés en milieu universitaire et par les organismes de développement (Trevisan et Walser 2014). Selon le tableau 1 qui suit, la conduite d'une étude d'évaluabilité peut se réaliser en suivant seulement quatre étapes, selon Dunn (2008), alors qu'on peut en compter jusqu'à dix si l'on considère la perspective de Smith (1989). La différence ne se situe pas seulement au niveau du nombre d'étapes à suivre, mais également dans les premières actions à poser. Ainsi, Wholey (1987) propose-t-il de commencer par impliquer les éventuelles parties prenantes au processus alors que, pour Thurston et Potvin (2003), il faut d'abord sélectionner la personne responsable de réaliser l'étude d'évaluabilité. Peu importe l'ordre des différentes étapes, Smith (1989; 1990) propose que les actions à mener soient guidées par le contexte et les objectifs poursuivis.

Tableau 1 : Les étapes de l'étude de l'évaluabilité de programmes selon quelques auteurs (*Adapté de Trevisan et Walser (2014) et Soura, Bastien et Fallu (2016). *Étapes ajoutées dans Wholey (2010).*)

Étapes	**Strosberg et Wholey (1983)**	**Wholey (1987, 2010)**	**Rog (1985)**	**Smith (1989)**
1	Déterminer les ressources, activités, objectifs et présuppositions du programme	Faire participer les éventuels utilisateurs et les parties prenantes	Analyser le programme	Déterminer le but, recueillir l'engagement, mettre en place le groupe de travail
2	Recueillir l'agrément des parties prenantes sur la description du programme	Clarifier la formulation du programme	Analyser l'implantation du programme	Définir les frontières du programme à évaluer
3	Déterminer les types de données à collecter	Explorer la réalité de l'implantation du programme	Analyser le système de mesure et d'information	Identifier et analyser les documents du programme
4	Vérifier la réalité de l'implantation du programme	Analyser la plausibilité du programme*	Effectuer une analyse de la plausibilité des objectifs du programme	Développer la théorie du programme
5	Déterminer si les ressources et activités mobilisées permettent d'atteindre les objectifs	S'entendre sur d'éventuels changements à apporter au programme ou à son implantation*	Préparer des modèles variés du programme	Identifier les parties prenantes et avoir des entrevues avec elles
6	Déterminer si le programme possède un système de gestion de l'information	S'entendre sur les options d'évaluation et les intentions d'utilisation des résultats	Déterminer l'utilisation des résultats de l'évaluation	Décrire la perception qu'ont les parties prenantes du programme
7	Déterminer la portion du programme prête à être évaluée en fonction des objectifs identifiés			Identifier les besoins, intérêts et différences de perceptions des parties prenantes

Étapes	**Strosberg et Wholey (1983)**	**Wholey (1987, 2010)**	**Rog (1985)**	**Smith (1989)**
8	Identifier les options d'évaluation et de gestion pour les gestionnaires du programme			Déterminer la plausibilité du modèle du programme
9				Tirer les conclusions et faire des recommandations
10				Établir un plan d'utilisation des données de l'étude d'évaluabilité

Tableau 1 (suite)

Étapes	**Thurston et Potvin (2003)**	**Dunn (2008)**	**Dunet et al. (2013)**
1	Sélectionner l'évaluateur responsable de conduire l'étude d'évaluabilité	Vérifier le modèle causal	Sélectionner le domaine de pratique
2	Identifier les parties prenantes du programme	S'accorder sur le but de l'évaluation d'impact	Identifier les programmes
3	Identifier et analyser la documentation	Évaluer la faisabilité d'un modèle alternatif	Rassembler les informations et évaluer l'intérêt de l'étude de l'évaluabilité
4	Développer le modèle logique du programme et le plan d'évaluation	Identifier une équipe d'évaluation locale	Sélectionner le(s) programmes à évaluer
5	S'accorder sur le principe de conduire une évaluation		Préparer les équipes pour les visites de terrain
6	Évaluer le temps et les ressources requis pour l'évaluation		Conduire les visites de terrain
7			Préparer les documents de l'étude de l'évaluabilité
8			Déterminer la possibilité d'une évaluation rigoureuse

Le premier modèle de la conduite d'une étude d'évaluabilité provient de Wholey (1987), l'un des pionniers de cette approche (Smith, 2005). Selon celui-ci, une étude d'évaluabilité devrait permettre d'apporter des réponses aux quatre conditions suivantes : 1) la théorie du programme et son modèle logique sont clairs, c'est-à-dire que les objectifs du programme et les informations prioritaires sont bien identifiées et définies; 2) les objectifs du programme sont plausibles; 3) les données pertinentes sur la performance du programme peuvent être obtenues à des coûts raisonnables; et 4) la manière d'utiliser les résultats de l'évaluation chez les utilisatrices et utilisateurs potentiels est clarifiée. En répondant à ces quatre conditions, l'étude d'évaluabilité devrait permettre, comme le souligne Smith (1989), de prendre une des cinq décisions suivantes : 1) arrêter le programme, 2) y

apporter des changements, 3) conduire une évaluation formelle, 4) laisser les choses dans leur état initial, ou 5) ignorer les résultats de cette étude.

L'étude d'évaluabilité peut se réaliser dans une période courte allant de quelques semaines à un ou deux mois, avant de conduire une évaluation formelle proprement dite pouvant se focaliser sur les processus et/ou les effets du programme. Elle peut également se prolonger dans le temps en fonction des opérations à mener et de la disponibilité des parties prenantes. Dans la plupart des études d'évaluabilité réalisées, les auteurs et les autrices ont le plus souvent recours aux données qualitatives collectées à l'aide d'entrevues, de revues documentaires, de groupes de discussion, de visites de terrain et d'observation (Esher *et al.* 2011). Les travaux sur l'étude d'évaluabilité menés avec des données quantitatives sont encore rares (Trevisan 2007). Toutefois, il faut retenir qu'au moment de la conduite de l'étude d'évaluabilité, la méthode sélectionnée sera guidée en grande partie par les impératifs logistiques, temporels et financiers.

Si, à première vue il peut sembler inutile d'entreprendre une étude d'évaluabilité avant une évaluation formelle, cette activité pré-évaluative est bien indiquée lorsque le programme à évaluer est qualifié de complexe (Contandriopoulos, Rey, Brousselle et Champagne 2012; Kania *et al.* 2013; Ogilvie *et al.* 2011; Rogers 2008; Touati et Suárez-Herrera 2012) et lorsque la théorie du programme et le modèle logique ne sont pas clairs ou encore lorsque le programme est peu documenté (Dunet *et al.* 2013; Rutman 1997). L'étude d'évaluabilité peut alors permettre de démêler l'écheveau de cette complexité en favorisant la description détaillée du programme et la mise en place d'outils nécessaires à une évaluation ultérieure.

L'étude d'évaluabilité peut être utile à la fois pour le programme, les évaluateurs et évaluatrices et les décideuses et décideurs (Soura, Dagenais *et al.*, 2016). On s'accorde aujourd'hui pour dire qu'entreprendre une étude d'évaluabilité peut être bénéfique à la phase d'élaboration du programme, lors de son implantation et pendant la phase post-implantation. En ce qui concerne les évaluateurs et évaluatrices, ils et elles peuvent tirer profit de l'étude d'évaluabilité en ce sens qu'elle est un moyen rapide d'appréhender l'état d'un programme et d'identifier les priorités des parties prenantes avant de déployer des moyens importants pour son évaluation. En outre, l'étude d'évaluabilité peut contribuer à une planification plus optimale des activités d'évaluation. Quant aux décideuses et décideurs, ils et elles ont là l'occasion de participer au processus d'évaluation en faisant valoir leur

point de vue et leurs intérêts. De plus, l'étude d'évaluabilité peut permettre aux gestionnaires d'opérer une meilleure distribution des ressources tant financières, matérielles qu'humaines en focalisant l'évaluation uniquement sur des aspects du programme susceptibles de répondre aux préoccupations des parties prenantes.

Exemple d'étude d'évaluabilité : l'intervention en réseau

Le programme à évaluer s'intitule l'*Intervention en réseau* (IR). Il s'agit d'un programme d'intervention destiné à prévenir l'usage de SPA en milieu scolaire. Il a été développé par la Commission scolaire des Navigateurs de Lévis au Québec à la suite d'une large consultation de l'ensemble des acteurs et actrices du milieu scolaire. Cette consultation avait mis en évidence l'importance de mettre en place des pratiques plus novatrices et adaptées de prévention de l'usage de substances psychoactives auprès des élèves. L'intervention développée devait mettre l'accent sur trois éléments fondamentaux, à savoir la création d'un lien significatif entre les élèves et la personne chargée de la mise en œuvre des activités, le démarrage des activités de prévention dès le primaire en suivant les élèves jusqu'au milieu du secondaire (3^{e} année) et la mise en place d'activités inspirées de l'Approche École en Santé (AES) (Désy 2009).

Une première version du projet a vu le jour à la rentrée scolaire de 2007-2008 avec le soutien financier du Fonds en partenariat du Centre de santé et des services sociaux (CSSS) du Grand-Littoral. Une étude de faisabilité a alors été entreprise en vue de valider et de mesurer le potentiel de cette pratique. Des aménagements ont été apportés afin d'aboutir à la forme actuelle de cette intervention. Pour la phase pilote, l'intervention a été déployée dans cinq écoles primaires et une école secondaire de ladite commission scolaire.

L'intervention en réseau (IR) fonctionne sur le modèle de l'intervenant-e pivot (IP) ou gestionnaire de cas. Ce modèle permet à un-e spécialiste d'assurer la coordination de plusieurs actions provenant de différentes entités afin de trouver une solution à une situation problématique (Farber, Deschamps et Cameron 2002; Initiative canadienne sur le cancer du sein 2002; Paradis, Racine et Gagné 2006). L'intervention consiste à placer un ou une IP dans une école secondaire rattachée à un groupe d'écoles primaires

situées dans la même zone géographique. Ce regroupement d'écoles secondaires et primaires est alors appelé *réseau*. L'IP peut être une personne déjà présente dans l'école à qui les tâches liées à l'intervention sont déléguées ou un-e professionnel-le recruté-e spécifiquement pour jouer ce rôle. Son lieu de présence habituelle est d'abord l'école secondaire. Il ou elle effectue ensuite des visites régulières au sein des écoles primaires qui lui sont attribuées. Ainsi, l'IP peut suivre les mêmes élèves sur une période plus ou moins longue si ces derniers restent plus longtemps dans l'école secondaire. Ce mode de fonctionnement lui permet de déployer des activités dans les deux ordres d'enseignement (primaire et secondaire) et favorise une meilleure connaissance des élèves. Ceux et celles à risque sont alors plus facilement repérées pour un suivi adéquat. Au-delà de cette présence physique, l'IP doit également mettre en œuvre des activités de prévention de l'usage de SPA avec la collaboration des enseignant-e-s et des autres intervenant-e-s des milieux scolaire et communautaire.

Pour les élèves du dernier cycle du primaire, les activités portent essentiellement sur les ateliers de *Système C et D*, deux programmes déjà utilisés dans les écoles au Québec. Ils ciblent particulièrement les facteurs de risque et de protection communs à la prévention de l'usage de SPA chez les jeunes en favorisant l'acquisition de saines habitudes de vie, la connaissance et l'affirmation de soi, la gestion du stress, la réaction face à l'influence des pairs et des médias et la résolution de conflits (Directions de santé publique des agences de la santé et des services sociaux de la Montérégie de Laval et de l'Outaouais, 2007). On y trouve également des activités sur la transition primaire-secondaire, par exemple les visites et rencontres des élèves du primaire avec ceux et celles du secondaire et les enseignant-e-s, la connaissance du fonctionnement de l'école secondaire, etc. Au secondaire (de la première à la troisième année), les élèves participent aux activités axées sur les ateliers du programme APTE (Garand-Butcher 2014; Vaugeois, Germain et Cunha Rêgo 2008). Il s'agit d'activités portant sur les croyances en matière de consommation de SPA, la connaissance des niveaux de risque, l'influence des pairs et les gangs de rue. Ces activités sont considérées comme un renforcement des acquisitions du primaire et sont animées par l'IP et d'autres intervenant-e-s qu'il ou elle sollicite (enseignant-e-s, psychoéducateurs et psychoéducatrices, infirmières et infirmiers, structures communautaires, le service de police, etc.). Les activités proposées aux élèves au primaire et au secondaire sont présentées au tableau 2.

Tableau 2 : Liste des activités de l'intervention en réseau

	Primaire (5e et 6e année)	Secondaire (1, 2 et 3)
1.	Saines habitudes de vie	Consommation de SPA et ses conséquences
2.	Connaissance de soi et affirmation de soi	Saines habitudes de vie
3.	Relations avec les autres et communication	Résistance aux influences négatives
4.	Influences, consommation et média	Relations enseignant-e-s-élèves
5.	Cyber criminalité	Violence et (cyber) criminalité
6.	Résolution de problèmes et gestion du stress	Comprendre le secondaire
7.	Préparation à l'entrée au secondaire	Rencontres individuelles
8.	Activités éducatives et récréatives	Activités éducatives et récréatives

Il est attendu qu'à court terme, l'intervention permette d'accroître les compétences sociales des élèves, d'améliorer les compétences en gestion du stress, d'améliorer les perceptions sur les risques de l'usage de substances psychoactives, de favoriser de bonnes relations entre enseignant-e-s et élèves, d'accroître les compétences en résolution de conflits et de réduire le niveau d'exposition des élèves aux SPA. À moyen terme, il s'agit de retarder l'âge d'initiation à l'usage de SPA et de prévenir la consommation excessive et problématique de ces substances. L'objectif ultime de cette intervention est de favoriser la réussite éducative des élèves.

Devis de l'étude d'évaluabilité

Pour réaliser cette étude d'évaluabilité, des données qualitatives ont été collectées essentiellement auprès d'informateurs-clés et d'informatrices-clés sélectionnées en raison de leur implication dans la conception et/ou la mise en œuvre de l'intervention. L'entrevue semi-structurée et l'analyse documentaire ont été les principales stratégies de collecte de ces données. L'objectif général était de bien connaître l'objet d'évaluation dans un premier temps avant d'amorcer son évaluation dans une phase ultérieure. Plus spécifiquement, il s'agissait de comprendre la théorie de l'intervention

compte tenu du fait que les rares documents disponibles à l'époque était vagues à ce sujet. À partir de cela, on pouvait élaborer le modèle logique et déterminer l'approche d'évaluation qui permettrait à la fois de comprendre le fonctionnement de l'intervention et ses potentiels effets sur la population cible.

S'appuyant sur les quatre conditions décrites par Wholey (1987) et présentées plus haut, cette étude d'évaluabilité visait à répondre aux questions suivantes :

- Les intentions des initiateurs et initiatrices et les objectifs de l'intervention sont-ils clairement présentés?
- Les objectifs de l'intervention sont-ils plausibles à la lumière des connaissances disponibles et des moyens mobilisés?
- Les conditions sont-elles réunies pour la collecte des données d'évaluation?
- Quelle sera l'utilité des résultats d'évaluation pour les parties prenantes?

Pour répondre à ces questions, une planification des activités a été réalisée. Celle-ci prend en compte les objectifs, les types de données collectées, les sources de données, les participant-e-s et la période de collecte. La collecte des données a été réalisée en un mois, une année après le démarrage de la mise en œuvre du programme, avec les participant-e-s comme informateurs et informatrices-clés. Cette planification est présentée dans le tableau 3.

Tableau 3. Planification des activités d'évaluation

Critères d'évaluation	Sujets/items abordés	Sources de données
Intentions et modèle logique de l'intervention	• les intentions et objectifs de l'intervention • les activités et ressources prévues et déployées • la population cible de l'intervention • les résultats attendus	• Entrevues • Revue documentaire
Plausibilité des objectifs	• les conditions d'implantation de l'intervention • les activités réalisées jusqu'à ce jour • l'adéquation des ressources avec les résultats attendus	• Entrevues • Revue documentaire • Journal de bord
Disponibilité des données	• l'identification d'une approche d'évaluation des effets et des processus • les types de données disponibles ou à recueillir • les activités de monitoring entreprises • la disponibilité de la documentation sur les activités	• Entrevues • Journal de bord • Notes de réunion
Utilité de l'intervention et des résultats de l'évaluation	• l'utilité de l'évaluation pour les initiateurs • les modalités d'utilisation des résultats de l'évaluation	• Entrevues • Notes de réunion

Les données colligées ont été analysées de façon thématique en quatre étapes. La première étape déductive a permis de générer des thèmes proches des conditions de l'approche de l'étude d'évaluabilité selon Wholey (1987). Une deuxième étape, cette fois-ci inductive, était ouverte à l'émergence de nouveaux thèmes (Saberi, Yuan, John, Sheon et Johnson 2013). Cette étape a permis d'identifier des thèmes faisant référence à la satisfaction des participant-e-s concernant la mise en œuvre du programme et la bonne collaboration avec l'IP. La troisième étape a consisté à regrouper tous les thèmes identifiés dans une matrice permettant d'identifier des sous-

catégories et des extraits *verbatim* pertinents au cadre d'analyse. Un réajustement des codes a été effectué afin de constituer de grands ensembles plus homogènes. Enfin la quatrième étape a permis de tirer des extraits de discours pour illustrer les résultats obtenus. Le traitement des données a été effectué à l'aide du logiciel de traitement de données qualitatives QDA Miner. Avant de présenter les résultats obtenus, les lignes qui suivent décrivent les différentes étapes de la mise en œuvre d'une étude d'évaluabilité.

Mise en œuvre de l'étude d'évaluabilité

La conduite d'une étude d'évaluabilité n'est pas une fin en soi. Cette activité doit répondre à un besoin particulier. Généralement elle sert de tremplin à une évaluation susceptible de produire des résultats utiles aux parties prenantes. En ce sens, elle peut permettre, comme déjà mentionné, de bien comprendre la théorie du programme et d'élaborer le modèle logique. Avant de réaliser une étude d'évaluabilité, il faudrait en expliquer le bien-fondé aux initiateurs et initiatrices du programme et à toutes les entités qui pourraient tirer profit des résultats obtenus. Les limites d'une telle activité devraient être aussi présentées afin de ne pas provoquer de vaines attentes. Une réunion de concertation et de clarification entre les différents acteurs et actrices peut être une excellente occasion de parvenir à cette fin.

Par ailleurs, étant donné que l'étude d'évaluabilité peut s'apparenter à une analyse du processus d'un programme afin de savoir si celui-ci répond à certains critères avant de l'évaluer de façon formelle, seuls les individus pouvant donner des renseignements sur ce programme doivent être contactés. Par conséquent, il faudrait s'assurer au préalable de la disponibilité de ces derniers et obtenir que les documents, s'il en existe, soient mis à la disposition de l'équipe d'évaluation. Finalement, bien connaître le contexte organisationnel et sociopolitique dans lequel se déroule le programme représente un atout considérable pour comprendre les jeux de pouvoir inhérents à la prise de décision; ce qui peut contribuer à la proposition de recommandations plus adaptées à ce contexte et susceptibles d'être adoptées par les parties prenantes.

Déroulement et description des étapes

Le déroulement d'une étude d'évaluabilité comprend plusieurs étapes qui ont été décrites dans la littérature. Plusieurs auteurs et autrices présentent différentes étapes à suivre pour mener une étude d'évaluabilité comme le montre le tableau 1 ci-dessus. Par souci pédagogique et pour répondre aux objectifs de cet ouvrage, la partie qui suit décrit les 10 étapes proposées par Smith (1989) dans l'approche qui est la plus exhaustive. Ces mêmes étapes ont été décrites de façon simplifiée ailleurs (Trevisan et Yi 2003). Il faut cependant noter qu'il ne s'agit pas d'un processus linéaire et il n'est pas exclu de revenir en arrière pour affiner les objectifs de l'étude d'évaluabilité alors que se déroulent les entrevues avec les parties prenantes, par exemple.

Déterminer le but, recueillir l'engagement, mettre en place le groupe de travail

Il est très important d'identifier le but à atteindre avant d'entreprendre une étude d'évaluabilité. Cela permet dès le départ de clarifier les attentes et de susciter l'engagement des parties prenantes. À cette étape également, mettre en place un groupe de travail peut favoriser l'engagement des parties prenantes et ainsi faciliter le déploiement des activités. Le groupe de travail mis en place doit être à la fois représentatif des différentes entités et restreint pour permettre son bon fonctionnement. Dans le cas présenté ici, le groupe de consultation était constitué principalement du concepteur de l'intervention, de l'intervenant pivot et de l'évaluateur principal. Les autres entités concernées ont été sollicitées lorsque le besoin se faisait sentir, notamment pour la validation du modèle logique.

Définir les limites/frontières du programme à évaluer

L'étude d'évaluabilité est destinée à être une activité exploratoire rapide de la situation d'un programme avant d'effectuer une évaluation proprement dite. À ce titre, elle ne peut apporter des réponses à tout. Il importe de fixer les limites quant à ce qui peut être réalisé ou pas dans la période impartie. Fixer des limites peut contribuer à mieux comprendre les objectifs assignés à l'étude d'évaluabilité et éviter que les uns et les autres aient des attentes que l'étude d'évaluabilité ne pourra combler. Dans le cas présenté ici, il était question ici de bien comprendre le programme, d'élaborer le modèle logique et de proposer des avenues pour l'évaluation du programme.

Identifier et analyser les documents liés au programme

La consultation des documents liés au programme est une étape importante de la conduite d'une étude d'évaluabilité. Il faut, au démarrage de l'étude d'évaluabilité, identifier les documents susceptibles de donner des renseignements utiles sur le programme et s'assurer de leur disponibilité. Il peut s'agir de notes de réunion, de rapport de recherche ou de sondage, de demande de financement, de documents législatifs et même de messages échangés entre les parties prenantes. Ces documents permettent de comprendre les intentions des initiateurs et initiatrices du programme et la théorie du changement promue. Il n'est pas rare, comme dans le cas présent, que l'on soit confronté à un nombre limité de documents qui aborde le programme. Dans ce cas, une consultation des personnes responsables de la mise en œuvre du programme sera nécessaire pour collecter des informations.

Développer/clarifier la théorie du programme

L'un des avantages de la conduite d'une étude d'évaluabilité est la possibilité de clarifier la théorie du programme et de développer le modèle logique. La théorie du programme concerne la manière dont ses initiateurs et initiatrices veulent s'y prendre pour atteindre les objectifs fixés. Elle explique comment les changements voulus par ces personnes seront obtenus à partir des ressources mobilisées, des stratégies adoptées et des activités menées. Dans de nombreuses situations, cette théorie n'est pas explicite ou est exprimée de façon vague ou parcellaire. Ce n'est qu'en parvenant à la démêler et à la clarifier que l'on parvient à bien comprendre le programme. Une fois la théorie du programme clarifiée, il est possible de développer le modèle logique qui est une représentation schématique de la chaîne causale présumée entre les ressources mobilisées, les activités déployées et les résultats attendus à moyen, court ou long terme. Il faudra également noter au passage les facteurs favorables ou défavorables qui pourraient influencer la mise en œuvre du programme dans un sens comme dans l'autre. Le développement du modèle logique doit être fait en concertation avec les parties prenantes. Dans tous les cas, la version finale ressemble rarement au modèle initial du fait de modifications successives qui peuvent être apportées au fur et à mesure que la théorie du programme se clarifie.

Identifier et interviewer les parties prenantes

L'identification et les entrevues avec les parties prenantes sont une autre étape du déroulement d'une étude d'évaluabilité, car ces personnes peuvent contribuer à une meilleure compréhension du programme. D'ailleurs, c'est grâce à leur soutien que le programme peut se déployer. Les entrevues doivent porter sur ce qu'elles savent et perçoivent du programme et de sa mise en œuvre. Dans le cas de l'IR, les personnes interrogées ont été identifiées avec la collaboration de l'IP qui était en contact direct avec elles. Une fois ces personnes identifiées, il est important de les contacter plusieurs jours d'avance afin de planifier l'organisation des entrevues.

Décrire la perception qu'ont les parties prenantes du programme

Les entrevues avec les parties prenantes doivent permettre de décrire la perception de ces informateurs et informatrices-clés à l'égard du programme. Au-delà de cette description, il faudra confronter les différents points de vue exprimés pour obtenir une meilleure compréhension des enjeux du programme en train d'être implanté ou déjà implanté.

Identifier les besoins, intérêts et différences de perceptions des parties prenantes

À ce stade, il appartient à l'évaluateur ou l'évaluatrice et à son équipe d'identifier les différences et les convergences au niveau des perceptions, besoins et intérêts des différents acteurs et actrices. Il faudra porter une attention particulière sur les divergences en raison de leurs potentiels effets néfastes sur la suite du programme. Comprendre les divergences et tenter de les résoudre permet d'éviter les malentendus et controverses qui peuvent naître lors de l'introduction des innovations dans les organisations.

Déterminer la plausibilité du modèle du programme

Lorsqu'elles sont bien conduites, l'analyse documentaire et les entrevues doivent permettre de déterminer la plausibilité du programme et de ses objectifs. Cette plausibilité est déterminée au regard des ressources mobilisées, mais aussi de la littérature existante. Il s'agit d'indiquer le degré avec lequel le programme est correctement et suffisamment mis en œuvre et si les activités telles que déployées sont susceptibles de provoquer les changements escomptés.

Tirer les conclusions et faire des recommandations

La conduite d'une étude d'évaluabilité doit aboutir à la formulation de recommandations susceptibles d'indiquer la conduite à adopter. Ces

conclusions et recommandations doivent être spécifiques au programme évalué et tirées uniquement à partir des données disponibles.

Établir un plan d'utilisation des données de l'étude d'évaluabilité

L'analyse des données recueillies doit permettre d'envisager la suite à réserver au programme. Les recommandations formulées doivent aller dans le sens de proposer des modifications au programme et/ou d'aller de l'avant avec une évaluation formelle ou pas. Il est également possible que les conclusions d'une étude d'évaluabilité préconisent d'arrêter le programme parce que la théorie du programme n'est pas adaptée au contexte ou à la population cible, faisant courir le risque d'effets iatrogènes chez ces dernières. L'arrêt du programme peut aussi s'envisager parce que l'étude d'évaluabilité a montré que sa mise en œuvre n'est pas conforme à ce qui a été planifié.

Résultats obtenus

Les résultats obtenus dans le cadre de l'étude d'évaluabilité de l'*Intervention en réseau* peuvent être regroupés en trois points. Commençons par le niveau de la clarification de la théorie du programme. Au début des activités d'évaluation de ce programme, il y avait très peu de documents sur l'approche adoptée par ses initiateurs et initiatrices. Si les objectifs à atteindre avaient été clairement présentés et qu'un intervenant avait été recruté pour la mise en œuvre du programme, la stratégie pour parvenir à l'atteinte de ces objectifs n'était pas aussi claire. De plus, certains acteurs et actrices sur le terrain avaient une idée vague de ce qui devait être réalisé. L'étude d'évaluabilité avait également montré que la mise en œuvre du programme était disproportionnée selon les écoles. À la suite de ces constats, des rencontres ont été organisées avec les différents acteurs et actrices pendant lesquelles une présentation plus approfondie du programme a été effectuée. Par ailleurs, des échanges ont eu lieu avec l'IP afin de lui présenter les résultats des consultations avec les autres acteurs et actrices et la nécessité de documenter les activités organisées. À la suite de cela, des dispositions ont été prises afin d'harmoniser le degré de mise en œuvre dans les écoles participantes. Par la même occasion, le suivi des activités a été nettement amélioré.

Continuons avec le développement du modèle logique. Une fois la

théorie du programme clarifiée, cela a rendu plus facile le développement du modèle logique qui n'existait pas au démarrage des activités d'évaluation. Celui-ci a été développé à l'issue de l'analyse de la documentation existante et des échanges avec les initiateurs et initiatrices du programme. La version initiale du modèle logique a subi plusieurs modifications avant que la version finale disponible ailleurs (Soura, Bastien, *et al.* 2016), ne soit adoptée pour servir de support à l'évaluation qui devrait être conduite.

Enfin, concernant la planification de l'évaluation, puisque l'équipe chargée de l'étude d'évaluabilité était la même qui devait conduire les activités d'évaluation proprement dites, des propositions ont été faites sur les modalités de réalisation de cette évaluation. Les échanges avec les parties prenantes et l'analyse des documents ont facilité l'identification de l'approche d'évaluation à adopter et le cadre d'analyse le plus adéquat pour rendre compte des processus de mise en œuvre du programme. En effet, il était apparu que la mise en œuvre du programme dans les écoles avait suscité des controverses chez certains acteurs et actrices. Par ailleurs, le programme avait un caractère novateur pour toutes les écoles impliquées. Tout cela a amené l'équipe d'évaluation à s'intéresser davantage à la façon dont les acteurs et actrices étaient mobilisés pour la mise en œuvre du programme qu'à l'analyse de la fidélité d'implantation. Ainsi, une revue de littérature a été réalisée dans le but d'identifier un cadre d'analyse pertinent qui tiendrait compte de cette situation. À la suite de cela, la théorie de l'acteur-réseau a été retenue pour servir de référence à l'évaluation des processus. Initialement développé pour l'étude de l'introduction d'innovations technologiques dans les organisations, ce cadre d'analyse s'est avéré pertinent par ce qu'il s'intéresse également aux controverses qui naissent lors de la mise en œuvre d'un programme.

Toujours concernant la planification de l'évaluation, la conduite de l'étude d'évaluabilité a également permis de confirmer l'adoption d'un devis quasi expérimental pour l'évaluation des effets du programme sur les participant-e-s. Ce devis qui avait été retenu avant le déroulement de l'étude d'évaluabilité a été préféré à une étude randomisée en raison des ressources disponibles et des contraintes liées à la mise en œuvre du programme. Il permet de comparer le groupe des élèves ayant participé au programme (groupe d'intervention) avec un autre groupe d'élèves qui n'y était pas exposé (groupe de comparaison). La faisabilité de ce volet quantitatif de l'évaluation a également été abordée lors de l'étude d'évaluabilité. En clair, la réalisation

de cette l'étude d'évaluabilité a grandement contribué à mieux comprendre le programme et à mettre en place les outils nécessaires pour l'évaluation qui s'en est suivie. Cette expérience nous a permis d'entreprendre une analyse réflexive sur la mise en œuvre d'une étude d'évaluabilité.

Analyse réflexive

Trois leçons principales peuvent être tirées de la conduite de cette étude d'évaluabilité. La première est que l'étude d'évaluabilité est un processus participatif qui demande une interaction constante entre l'équipe d'évaluation et les parties prenantes. Même s'il n'est pas possible d'avoir la coopération entière de ces dernières, il est cependant indispensable et très important, au moins, qu'un processus de validation des résultats soit mis en place afin de recueillir leurs avis. Cela a pour bénéfice de maximiser les chances de l'adoption des recommandations, notamment lorsque celles-ci préconisent des modifications majeures dans la mise en œuvre du programme ou une réorientation de ses objectifs. Au surplus, faire participer les parties prenantes à la réalisation de l'étude d'évaluabilité permet de mieux documenter le programme et de proposer des recommandations pertinentes à celui-ci.

Deuxièmement, la conduite d'une étude d'évaluabilité est bénéfique non seulement pour l'équipe d'évaluation, mais également pour les intervenant-e-s eux-mêmes qui ont l'occasion d'avoir une attitude réflexive, ce qui peut contribuer à l'amélioration de leurs pratiques. Les étapes de la réalisation d'une étude d'évaluabilité et les *checklists* développées par les différents auteurs et autrices sont de bons moyens pour commencer si l'on ne sait par où commencer. Ces outils doivent cependant être utilisés de façon flexible et surtout être adaptés aux réalités et au contexte du programme à évaluer ainsi qu'à l'audience.

Troisièmement, même si les ressources sont limitées, conduire une étude d'évaluabilité, ne serait-ce qu'à une échelle plus réduite, peut être très bénéfique aussi bien pour le programme et ses initiateurs et initiatrices que pour l'équipe d'évaluation. Cette étude d'évaluabilité a été pour toutes les parties prenantes un excellent moyen de mieux se connaître en prélude aux activités d'évaluation. Par ailleurs, en encourageant la participation des parties prenantes, l'étude d'évaluabilité favorise l'appropriation du processus

par les parties prenantes et représente à ce titre un excellent exercice d'*empowerment* de ces entités qui n'ont pas toujours de compétences en matière d'évaluation de programmes. Si le temps l'avait permis et si les conditions de travail avaient été différentes, une étude plus approfondie aurait peut-être permis une meilleure planification de cette évaluation afin de raccourcir la durée de celle-ci.

Comme indiqué dans l'encadré ci-dessus, l'étude d'évaluabilité possède des points forts, mais également quelques points faibles. Comme points forts, notons que l'étude d'évaluabilité encourage un processus participatif pendant lequel toutes les entités concernées par le programme et l'évaluation interagissent. La prise en compte des points de vue de différentes entités permet ainsi d'améliorer le programme et de proposer des résultats d'évaluation adaptés aux préoccupations des utilisateurs. L'étude d'évaluabilité permet également de clarifier la théorie du programme et d'élaborer le modèle logique. En effet, grâce à l'étude d'évaluabilité, l'on peut parvenir à une meilleure compréhension de la théorie d'un programme et des réelles intentions de ses initiateurs et initiatrices, ce qui peut favoriser le développement du modèle logique qui reste un outil indispensable pour l'évaluation d'un programme. Un autre avantage à conduire une étude d'évaluabilité est que celle-ci peut contribuer à apporter des améliorations substantielles au programme que ce soit à la phase de la conception ou pendant l'implantation. C'est un exercice qui permet de sauver du temps et des ressources puisqu'elle peut recommander de ne pas engager une évaluation si celle-ci n'est d'aucune utilité pour les parties prenantes. Enfin, l'étude d'évaluabilité peut faciliter la planification de l'évaluation à venir en proposant des stratégies et l'identification d'approches et d'outils pertinents au contexte de mise en œuvre du programme et à la genèse de connaissances utiles à toutes les parties prenantes.

Le principal point faible de l'étude d'évaluabilité concerne le fait qu'étant donné qu'elle requiert un processus participatif, elle peut également prendre beaucoup de temps à réaliser si le groupe de travail mis en place ne fonctionne pas de façon optimale ou s'il n'y a pas d'entente entre ses membres ou encore si le nombre de personnes à consulter est élevé. L'autre point faible est que l'étude d'évaluabilité peut également exacerber les mésententes et provoquer des frustrations chez certains acteurs et actrices, notamment lorsque les attentes des uns et des autres ne sont pas assouvies. Il est donc très important de s'accorder sur les objectifs à atteindre dès le

départ de ce processus. Il va donc sans dire que la mise en œuvre d'une étude d'évaluabilité présente des défis dépendamment du contexte dans lequel celle-ci se déploie.

Le premier défi lors de la réalisation de cette étude d'évaluabilité était celui de pouvoir mobiliser toutes les parties prenantes à participer effectivement au processus compte tenu des contraintes de temps liées aux activités scolaires. Le programme se déroulant dans les écoles, il n'était pas toujours évident pour les enseignant-e-s et les autres professionnel-le-s de se libérer en vue de participer aux entrevues. Trouver les ressources humaines et financières nécessaires à la conduite de l'étude d'évaluabilité alors que celle-ci n'était pas inscrite parmi les objectifs à atteindre au début du mandat d'évaluation était un autre défi de ce processus. Cette situation combinée à la longue distance entre le site d'implantation du programme et le lieu d'exercice habituel des évaluateurs et évaluatrices était de nature à retarder la mise en œuvre de l'étude d'évaluabilité dans les délais impartis. Lorsque l'étude d'évaluabilité n'est pas une des préoccupations des parties prenantes, il peut être difficile de leur demander d'attendre que les résultats de celle-ci soient disponibles avant de commencer les activités d'évaluation proprement dites. Enfin, sans avoir été forcément un défi dans la conduite de la présente étude d'évaluabilité, mais qui mérite d'être abordés ici, ce sont les enjeux de pouvoir : comment concilier les points de vue des différents acteurs et actrices, en particulier lorsqu'ils sont divergents. C'est pourquoi les qualités de facilitateur de groupe de travail et de négociateur sont des caractéristiques essentielles qu'une personne doit posséder pour bien mener à terme une étude d'évaluabilité.

Conclusion

En résumé, notons que l'étude d'évaluabilité a gagné en importance en termes d'utilisation ces dernières années. En effet, depuis son développement, de nombreux travaux ont contribué à faciliter son application en proposant les étapes à suivre et des *checklists* comme guides. Cependant, son usage dans le milieu francophone reste encore limité. Pourtant, conduire une étude d'évaluabilité est un intéressant moyen de parvenir à une bonne connaissance du programme à évaluer et au développement d'outils nécessaires à la planification des activités

d'évaluation à venir. Elle a été d'une grande utilité à plusieurs égards dans le cadre de l'évaluation du programme utilisé ici en exemple, d'abord en favorisant une plus grande familiarité entre les parties prenantes et le programme, ensuite en permettant de développer le modèle logique et enfin en aidant à la planification des activités d'évaluation. Les résultats obtenus incitent à penser qu'il aurait été difficile de procéder autrement pour un programme dont la mise en place constituait un défi pour les initiateurs et initiatrices.

Pour les novices à cette approche, il est possible de recourir aux étapes et *checklists* proposées par les auteurs et autrices. Toutefois, il faut garder à l'esprit que la conduite d'une étude d'évaluabilité est influencée par les caractéristiques du programme, celles des parties prenantes impliquées dans le processus ainsi que par le contexte d'implantation. C'est pourquoi il peut être nécessaire d'adapter le processus en fonction de ces réalités. L'étude d'évaluabilité ne pourra peut-être pas répondre à toutes les questions, mais elle peut être une alternative utile pour éviter le *statu quo* ou la conduite d'une évaluation dont les résultats ne peuvent être utilisés par aucune des parties.

Références clés

Lessard, S., Stich, C. et Montreuil, A. (2012). *Évaluation de l'évaluabilité du Plan québécois de prévention du tabagisme chez les jeunes 2010-2015*. Québec : Institut National de Santé Publique du Québec.

Ce rapport de recherche décrit la conduite de l'étude d'évaluabilité du Plan québécois de prévention du tabagisme chez les jeunes 2010-2015. Il présente d'abord le mandat confié à l'équipe d'évaluation ainsi que la méthode adoptée et les résultats obtenus. C'est un excellent document qui montre le processus de validation du modèle logique auprès d'acteurs et actrices-clés.

Burrows, S., Bilodeau, A. et Litvak, E. (2012). *Étude de la faisabilité de l'évaluation et des possibilités de recherche du Plan d'action montréalais 2012-2014 de la table régionale sur les saines habitudes de vie et la prévention des problèmes liés au poids – Phase 1*. Montréal : Direction de santé publique, Agence de la Santé et des Services Sociaux de Montréal.

Ce rapport porte sur une étude d'évaluabilité réalisée pour déterminer le potentiel de recherche et/ou d'évaluation de plusieurs projets entrant dans

le cadre du Plan d'action montréalais de la table régionale sur les saines habitudes de vie et la prévention des problèmes liés au poids. Le document donne un aperçu de la façon dont on peut réaliser une étude d'évaluabilité pour plusieurs projets à la fois.

Trevisan, M. S. et Walser, T. M. (2014). *Evaluability assessment : Improving evaluation quality and use*. Thousand Oaks: SAGE Publications, Inc.

Cet ouvrage en anglais est l'un des plus récents et par conséquent l'un des plus actuels sur l'étude d'évaluabilité. Il est présenté par deux des auteurs les plus prolixes sur cette approche. Ce qui le rend intéressant à consulter, c'est la référence à des études de cas qui permettent de comprendre comment conduire une étude d'évaluabilité. Par ailleurs, les auteurs militent pour un modèle de la conduite d'une étude d'évaluabilité à quatre étapes et introduisent dans cet ouvrage d'autres concepts fondamentaux à l'évaluation comme le développement de programmes, l'évaluation formative et des impacts.

Site Monitoring and Evaluation NEWS
http://mande.co.uk/category/lists/evaluability-assessments-bibliography/
https://www.zotero.org/groups/211251/evaluability_assessments/items

Ce site Internet est le tout premier à répertorier des ouvrages datant des premiers moments de développement de l'étude d'évaluabilité, depuis les années 1970 jusqu'en 2012. Ce répertoire est actuellement en voie d'être actualisé par son auteur et est consultable sur la plateforme de gestion bibliographique Zotero, à partir du lien ci-dessus. Bien que la plupart des documents soient en anglais, c'est une excellente ressource pour quiconque voulant localiser des références sur l'étude d'évaluabilité.

http://aea365.org/blog/tag/evaluability-assessment/

Ce site est un blogue de l'Association Américaine d'Évaluation spécifiquement dédié à l'étude d'évaluabilité. Il s'agit de billets très courts postés par des praticiens ou chercheuses et chercheurs sur leurs expériences dans la conduite d'une étude d'évaluabilité. Les auteurs et les autrices font référence aux leçons apprises de ces expériences, ce qui peut être très utile pour les personnes voulant conduire une étude d'évaluabilité pour la première fois.

Références

Burrows, S., Bilodeau, A. et Litvak, E. (2012). *Étude de la faisabilité de l'évaluation et des possibilités de recherche du Plan d'action montréalais 2012-2014 de la table régionale sur les saines habitudes de vie et la prévention des problèmes liés au poids – Phase 1*. Montréal : Direction de santé publique de l'Agence de la santé et des services sociaux de Montréal.

Contandriopoulos, A. P., Rey, L., Brousselle, A. et Champagne, F. (2012). Évaluer une intervention complexe : enjeux conceptuels, méthodologiques et opérationnels. *La Revue Canadienne d'Évaluation de Programmes*, 26(3), 1-16.

Craig, P. et Campbell, M. (2015). *Evaluability assessment: a systematic approach to deciding whether and how to evaluate programmes and policies*. Scotland : What Works Scotland.
http://whatworksscotland.ac.uk/wp-content/uploads/2015/07/WWS-Evaluability-Assessment-Working-paper-final-June-2015.pdf

Davies, R. (2013). *Planning evaluability assessments: A synthesis of the literature with recommendations*.UK : Department of International Development.

Désy, M. (2009). *L'école en santé : recension des écrits*. Montréal: Direction de santé publique de l'Agence de la santé et des services sociaux de Montréal.

Directions de santé publique des agences de la santé et des services sociaux de la Montérégie de Laval et de l'Outaouais. (2007). *Système D : activités de développement des compétences en milieu scolaire pour la prévention des toxicomanies et de l'usage des jeux de hasard et d'argent dans le cadre de la transition du primaire au secondaire*. Longueuil, QC.

Dunet, D. O., Losby, J. L. et Tucker-Brown, A. (2013). Using evaluability assessment to support the development of practice-based evidence in public health. *Journal of Public Health Management and Practice*, 19(5), 479-482.
doi: 10.1097/Phh.0b013e318280014f

Dunn, E. (2008). *Planning for cost effective evaluation with evaluability assessment*. Washington, DC: USAID.
http://pdf.usaid.gov/pdf_docs/PNADN200.pdf

Esher, A., dos Santos, E. M., Azeredo, T. B., Luiza, V. L., Osorio-de-Castro, C.

G. S. et Oliveira, M. A. (2011). Logic models from an evaluability assessment of pharmaceutical services for people living with HIV/AIDS. *Ciencia et Saude Coletiva*, 16(12), 4833-4844.

Farber, J. M., Deschamps, M. et Cameron, R. (2002). *Analyse et évaluation de la fonction d'intervenant-pivot en réponse aux besoins qu'éprouvent les Canadiennes atteintes d'un cancer du sein sur les plans de l'information, de l'éducation et de l'aide à la prise de décisions*. Gouvernement du Canada. http://publications.gc.ca/collections/Collection/H39-663-2002F.pdf

Garand-Butcher, C. (2014). *Évaluation de l'implantation du programme Système D en Montérégie* (Mémoire de maîtrise). Université de Montréal, Montréal, QC.

Horst, P., Nay, J. N., Scanlon, J. W. et Wholey, J. S. (1974). Program management and the federal evaluator. *Public Administration Review*, 34(4), 300-308.
doi:10.2307/975239

Initiative Canadienne sur le Cancer du Sein. (2002). *Analyse et évaluation de la fonction d'intervenant-pivot en réponse aux besoins qu'éprouvent les Canadiennes atteintes d'un cancer du sein sur les plans de l'information, de l'éducation et de l'aide à la prise de décisions*. Canada.

Kania, A., Patel, A. B., Roy, A., Yelland, G. S., Nguyen, D. T. K. et Verhoef, M. J. (2013). Capturing the complexity of evaluations of health promotion interventions: A scoping review. *The Canadian Journal of Program Evaluation*, 27(1), 65-91.

Kaufman-Levy, D. et Poulin, M. (2003). *Evaluability assessment: Examining the readiness of a program for evaluation*. Justice Research and Statistics Association, United States of America.
https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=202398

Leviton, L. C., Khan, L. K., Rog, D., Dawkins, N. et Cotton, D. (2010a). Evaluability assessment to improve public health policies, programs, and practices. *Annual Review of Public Health*, 31, 213-233.
doi: 10.1146/annurev.publhealth.012809.103625

Nay, J. N. et Kay, P. (1982). *Government oversight and evaluability assessment: It is always more expensive when the carpenter types*. Lexington, MA: Lexington.

Ogilvie, D., Cummins, S., Petticrew, M., White, M., Jones, A. et Wheeler, K.

(2011). Assessing the evaluability of complex public health interventions: Five questions for researchers, funders, and policymakers. *The Milbank Quarterly*, 89(2), 206-225.
doi: 10.1111/j.1468-0009.2011.00626.x

Paradis, M., Racine, P. et Gagné, D. (2006). *Évaluation des interventions et des activités liées à la problématique de l'abus, la négligence et la violence à l'endroit des aînés*. Québec: Direction régionale de santé publique de l'Agence de la santé et des services sociaux de la Capitale-Nationale.

Rog, D. J. (1985). A *methodological analysis of evaluability assessment* (Thèse de doctorat). Université Vanderbilt, Nashville, TN.

Rogers, P. J. (2008). Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation*, 14(1), 29-48.

Rutman, L. (1997). Formative research and program evaluability. Dans L. Rutman (dir.), *Evaluation research methods*: A *basic guide* (p. 59-71). Beverly Hills, CA : Sage Publications.

Saberi, P., Yuan, P., John, M., Sheon, N. et Johnson, M. O. (2013). A pilot study to engage and counsel HIV-positive African American youth via telehealth technology. *AIDS patient care and STDs*, 27(9), 529-532.

Smith, M. F. (1989). *Evaluability assessment: A practical approach*. Boston Norwell, Mass. : Kluwer Academic; Distributors for North America Kluwer Academic Publishers.

Smith, M. F. (1990). Evaluability assessment: Reflections on the process. *Evaluation and Program Planning*, 13(4), 359-364.
doi: 10.1016/0149-7189(90)90020-W

Smith, M. F. (2005). Evaluability assessment. Dans S. Mathison (dir.), *Encyclopedia of evaluation* (p. 137-140). Thousand Oaks, CA: Sage Publications.

Soura, B. D., Bastien, R. et Fallu, J.-S. (2016). Étude d'évaluabilité d'une intervention visant à prévenir l'usage de substances psychoactives lors de la transition primaire-secondaire. *Revue Canadienne d'Évaluation de Programme*, 31(2), 211-231.

Soura, B. D., Dagenais, C., Bastien, R., Fallu, J.-S. et Janosz, M. (2016). L'étude d'évaluabilité : pertinence pour l'évaluation de programme. *Revue Canadienne d'Évaluation de Programme*, 31(1), 18-33.

Thurston, W. E. et Potvin, L. (2003). Evaluability assessment: A tool for incorporating evaluation in social change programmes. *Evaluation*, 9(4), 453-469.

Touati, N. et Suárez-Herrera, J. C. (2012). L'évaluation des interventions complexes : quelle peut être la contribution des approches configurationnelles?. *La Revue Canadienne d'Évaluation de Programmes*, 26(3), 17-35.

Trevisan, M.S. (2007). Evaluability assessment from 1986 to 2006. *American Journal of Evaluation*, 28(3), 290-303. http://doi:10.1177/1098214007304589

Trevisan, M. S. et Walser, T. M. (2014). *Evaluability assessment : Improving evaluation quality and use*. Thousand Oaks, CA: Sage Publications.

Trevisan, M. S. et Yi, M. H. (2003). Evaluability assessment: A primer. *Practical Assessment Research et Evaluation*, 8(20), 2-9.

Van Voorhis, P. V. et Brown, K. (1987). *Evaluability assessment: A tool for program development in corrections*. National Institute of Corrections.

Vaugeois, P., Germain, M. et Cunha Rêgo, L. (2008). *Programme APTE (Rapport synthèse)*. Montréal : Centre québécois de lutte aux dépendances.

Wholey, J. S. (1976). Role of evaluation and evaluator in improving public programs: Bad news, good news, and a bicentennial challenge. *Public Administration Review*, 36(6), 679-683. doi: 10.2307/975063

Wholey, J. S. (1987). Evaluability assessment: Developing program theory. *New Directions for Program Evaluation*, (33), 77-92. doi: 10.1002/ev.1447

Wholey, J. S. (2010). Exploratory evaluation. Dans J. S. Wholey, H. P. Hatry et K. E. Newcomer (dir.), *Handbook of practical program evaluation* (p. 81-99). San Francisco, CA: Jossey-Bass.

Résumé / Abstract / Resumen

Réaliser une étude d'évaluabilité (ÉÉ) est une étape importante permettant de planifier l'évaluation proprement dite d'une intervention ou d'un programme. C'est un exercice qui permet de collecter rapidement des

données sur le déroulement d'un programme afin de prendre des décisions sur d'éventuelles modifications à apporter ou la conduite d'une évaluation formelle. Dans le cas actuel, l'étude d'évaluabilité a été appliquée à un programme de prévention de l'usage de substances psychoactives en milieu scolaire. Les différentes étapes de la conduite d'une étude d'évaluabilité y sont décrites ainsi que les résultats obtenus et les leçons apprises. Par ailleurs, les forces et faiblesses de l'étude d'évaluabilité, en tant que phase pré-évaluative d'un programme, et les défis de sa mise en œuvre sont présentés.

***

Conducting an evaluability Assessment (EA) is an important step in planning the actual evaluation of an intervention or program. It is an exercise that allows to quickly collect data on the implementation of a program in order to make decisions about possible changes or conduct a formal evaluation. In this actual case, EA was applied to a school-based substance abuse prevention program. The different steps to conduct an EA are described as well as the results and the lessons learned. In addition, the strengths and weaknesses of EA, as a pre-evaluative phase of a program, and the challenges of its implementation are presented.

***

La realización de una Evaluación de la Evaluabilidad (EE) es un paso importante en la planificación de la evaluación real de una intervención o programa. Es un ejercicio que permite recopilar datos rápidamente sobre el progreso de un programa con el fin de tomar decisiones sobre posibles cambios o realizar una evaluación formal. En el caso presente, la EE se aplicó a un programa de prevención de abuso de sustancias basado en la escuela. Se describen las diferentes etapas de la conducción de una EE, los resultados obtenidos y las lecciones aprendidas. Además, se presentan las fortalezas y debilidades de la EE, como una fase previa a la evaluación de un programa, y los desafíos de su implementación.

***

Biessé D. Soura est actuellement en stage postdoctoral aux *Centers for Disease Control and Prevention (CDC)* à Atlanta au sein du service de suivi, évaluation et analyse des données. Dans le cadre de ses activités, il effectue des missions dans les pays recevant les fonds d'urgence présidentiels pour la lutte contre le VIH du gouvernement américain tels que le Rwanda, le

Cameroun, la Côte d'Ivoire et Haïti. Il est un ex-boursier du Programme de formation Prévention, Promotion et Politiques Publiques (4P) du Réseau de Recherche en Santé des Populations du Québec (RRSPQ). Il détient un doctorat en psychoéducation (Université de Montréal), une maîtrise en psychologie des organisations et du travail (Université Paris Descartes) et un certificat en santé publique (*Emory University, Atlanta*). Sa thèse porte sur l'évaluation des processus et des effets d'une intervention visant à prévenir l'usage de substances psychoactives (SPA) en milieu scolaire et a reçu l'un des prix de la meilleure thèse du RRSPQ pour l'année 2016. Il s'intéresse à l'utilisation d'approches innovantes et participatives pour le suivi et l'évaluation de programmes, la prévention du VIH et de la consommation de SPA, les méthodes de recherche mixtes et le transfert des connaissances.

Jean-Sébastien Fallu est professeur agrégé à l'École de psychoéducation de l'Université de Montréal. Il détient un doctorat en psychologie de la même université et a réalisé un stage postdoctoral au Centre de toxicomanie et de santé mentale, affilié à l'Université de Toronto. Il est chercheur régulier à l'Institut universitaire sur les dépendances du Centre intégré universitaire de santé et de services sociaux du Centre-Sud-de-l'Île-de-Montréal, au Groupe de recherche sur les substances psychoactives et à l'Institut de recherche en santé publique de l'Université de Montréal. Il œuvre par ailleurs en tant que Directeur de la revue *Drogues, santé et société*. Ses intérêts de recherche portent notamment sur l'étiologie, la prévention et la réduction des méfaits de la consommation problématique de substances ainsi que sur les politiques en la matière. Fondateur du Groupe de recherche et d'intervention psychosociale de Montréal, cette réalisation lui a valu le prix Forces Avenir 2002, dans la catégorie *Société, communication et éducation*.

Robert Bastien est chercheur à la Direction de la santé publique (DSP) de Montréal et professeur adjoint de clinique au Département de médecine sociale et préventive de l'Université de Montréal. Il détient une maîtrise et un doctorat en éducation, ainsi qu'un baccalauréat en design. Il s'intéresse aux politiques, aux discours et aux pratiques de prévention. Ses travaux tentent de cerner les vecteurs d'influence de la recherche sociale sur les politiques et d'identifier de nouveaux leviers pour que la recherche contribue à renouveler les politiques, les pratiques et les services. En collaborant avec des personnes œuvrant dans le domaine des arts, du social et de la santé, il provoque la rencontre entre diverses disciplines, développe de nouveaux

questionnements, et expérimente de nouvelles méthodologies de recherche et de création.

Frédéric N. Brière est professeur adjoint à l'École de Psychoéducation de l'Université de Montréal et chercheur régulier du Groupe de Recherche sur les Environnements Scolaires (GRES), de l'Équipe RENARD et de l'Institut de Recherche en Santé Publique de l'Université de Montréal (IRSPUM). Il s'intéresse depuis des années au développement, aux conséquences et à la prévention de la dépression chez les adolescents. De manière plus générale, ses travaux portent sur l'évaluation de programmes visant à favoriser le bon développement psychosocial des jeunes et les meilleures façons de transférer ces interventions fondées sur les données probantes de la recherche vers la pratique.

Citation

Biessé Diakaridja Soura, Jean-Sébastien Fallu, Robert Bastien et Frédéric N. Brière. (2019). L'étude d'évaluabilité. Une intervention de prévention de l'usage de drogues à l'école au Québec. In *Évaluation des interventions de santé mondiale. Méthodes avancées*. Sous la direction de Valéry Ridde et Christian Dagenais, pp. 3-32. Québec : Éditions science et bien commun et Marseille : IRD Éditions.

2. L'évaluation de la pérennité

Une intervention de financement basé sur les résultats au Mali

MATHIEU SEPPEY ET VALÉRY RIDDE

L'évaluation de la pérennité

Définition de la méthode

La pérennisation est le processus qui conduit à la pérennité d'une intervention. L'évaluation de la pérennité est l'étude de la présence de routines pour l'attribution d'un degré de pérennité (nul, précaire, faible, moyen, haut). L'évaluation de la pérennisation consiste à analyser les processus qui ont permis à l'intervention de perdurer dans le temps et d'expliquer le degré de pérennité.

Forces de la méthode

- Mieux comprendre le processus de pérennisation, un phénomène complexe dans un contexte spécifique, à étudier selon de multiples dimensions
- Développer une analyse basée sur un ensemble exhaustif de points de vue
- Trianguler des méthodes de collecte de données (entrevues, observations, recherche documentaire)

Défis de la méthode

- Besoin de ressources (temps, budget) pour analyser les données
- Tenir compte de la validité interne de l'information (plusieurs biais possibles)
- Difficulté à appréhender le concept et à faire le va-et-vient entre la théorie et les données de terrain

Le présent chapitre présente l'évaluation de la pérennité à travers l'exemple de l'évaluation d'une intervention en santé dans divers centres de santé au Mali, plus précisément un projet pilote de financement basé sur les résultats (FBR).

Parmi les différentes méthodologies possibles pour évaluer la pérennité d'une intervention, celle utilisée ici est l'étude de cas (Yin 2013) qui peut se décliner en cas unique, cas multiple, cas à niveaux d'analyse imbriqués ou holistique, etc. L'étude présentée ici est celle d'un cas unique avec différents niveaux d'analyse imbriqués, à savoir différents centres de santé offrant divers services.

Plusieurs évaluations de la pérennité se limitent à étudier l'aspect financier de l'intervention ou ses effets à long et moyen termes. Utiliser l'étude de cas permet d'investiguer davantage de dimensions de l'intervention évaluée : ses ressources, son fonctionnement, ses déterminants (facilitateurs ou barrières), ses effets à court et long termes, etc. On investiguera donc ces dimensions à travers différents instruments de collecte de données (entretiens, observations, recherches documentaires, groupes de discussion, etc.), ce qui permettra de trianguler l'information recueillie et ainsi de mieux comprendre le phénomène global à l'étude. À noter qu'en complément de l'étude de cas, l'évaluation rapportée ici s'est aussi servie d'un cadre conceptuel de la pérennité provenant de la littérature scientifique. Ce cadre permet d'adopter un point de vue systémique sur le phénomène à l'étude et de mieux structurer l'évaluation en mettant l'accent sur les dimensions essentielles et sous-jacentes au phénomène.

La méthodologie de l'étude de cas est très utile dans l'évaluation de phénomènes complexes comme la pérennité puisqu'elle permet l'obtention d'un maximum d'informations sur une variété de dimensions liées à l'objet d'étude (Contandriopoulos *et al.* 1990; Lamoureux 2000; Yin 2013). Pour Robert Yin (2013), ce type de démarche méthodologique devrait toutefois se restreindre à trois situations : 1) lors d'un besoin de réponse à des questions de recherche telles que « Comment? » ou « Pourquoi? », 2) lorsque que très peu de contrôle ou aucun contrôle ne peut être effectué par le chercheur ou la chercheuse sur l'intervention et 3) quand le sujet de l'étude est contemporain, ce qui permet d'explorer directement le sujet à l'étude. Cette méthodologie est aussi nécessaire lorsque les acteurs et actrices concernés par l'évaluation sont nombreux, diversifiés et portent différents points de vue. L'utilisation d'un cadre conceptuel à l'intérieur d'une étude de cas permet aussi de faciliter la compréhension de l'objet de l'étude en identifiant les différents éléments reliés à ce dernier.

Plusieurs défis d'application sont toutefois à noter. Le premier est relié à la quantité d'informations nécessaires à la compréhension du cas. Plus l'intervention à évaluer est complexe, plus le chercheur ou la chercheuse aura besoin d'informations pour la comprendre. De plus, pour certains détails de l'intervention, des informations devront être recueillies auprès de personnes-ressources ou d'expert-e-s spécifiques. Le deuxième défi est lié à la complexité de l'analyse de ces données qui proviennent d'une grande diversité de sources (documents, entrevues, observations, etc.) et d'acteurs

et actrices. Insérer toutes ces informations à l'intérieur d'une même explication peut s'avérer difficile. Plusieurs outils et logiciels informatiques sont toutefois utilisés aujourd'hui pour aider les chercheuses et chercheurs à mener ce type d'analyse.

Contexte

Le Mali

L'étude de cas présentée dans ce chapitre porte sur un projet pilote de financement basé sur les résultats (FBR) se déroulant au Mali. Le Mali est un pays sahélien enclavé d'Afrique de l'Ouest se classant au 179^{e} rang de l'index de développement humain en 2015 (United Nations Development Programme 2015). En 2015, le pays était constitué de plus de 15,8 millions d'habitants, dont la moitié (50,6%) vivait sous le seuil de la pauvreté avec 1,25 $US/jour en parité du pouvoir d'achat (*ibid.*). En 2015, la mortalité néonatale ainsi que la mortalité avant cinq ans étaient très élevées avec respectivement 77,6 et 122,7 décès pour 1 000 naissances vivantes. En 2014, le total des dépenses en santé n'était que de 6,9% du PIB (United Nations Statistics Division 2016).

Le Mali est un pays dont l'administration est très décentralisée, avec plusieurs niveaux de représentation : national, régional, par district, par cercle et communal; les villages et quartiers pourraient aussi être ajoutés, mais n'ont pas de compétences dans les domaines de la santé et de l'éducation (Syll 2005). L'organisation du système de santé calqué sur ce modèle comporte des hôpitaux nationaux, spécialisés et régionaux qui sont dirigés par la Direction nationale de la santé (DNS). Ces institutions de santé représentent le niveau tertiaire (de 2^{e} ou 3^{e} référence) de prestations de soins et se distribuent dans les huit régions maliennes. Les centres de première référence (niveau secondaire) sont les centres de santé de référence (CSREF) qui sont présents dans les districts sanitaires reliés au niveau administratif des cercles (à ne pas confondre avec les districts bamakois) (Observatoire africain de la santé 2014). Dans les services de première ligne, on retrouve les centres de santé communautaire (CSCOM), qui se situent dans les aires de santé liées aux communes. La gestion quotidienne des CSCOM est particulière puisqu'elle est prise en charge par

une association de santé communautaire (ASACO) formée d'élus et d'élues des différents villages de la commune et qui s'assure du bon fonctionnement du CSCOM et de la provision du paquet minimum d'activités (PMA) (Observatoire africain de la santé 2014; Audibert et de Roodenbeke 2005). Le PMA représente tous les services qu'un CSCOM devrait offrir, soit des services promotionnels, préventifs et curatifs. Les activités vont du simple diagnostic et traitement d'une maladie à des campagnes de vaccinations, des suivis d'enfants sains (SES), des séances d'IEC (information, éducation, communication), des consultations pré/postnatales (CPN et CPON), des accouchements, des planifications familiales (PF) et autres. Ces divers services sont offerts par les trois composantes du CSCOM : le dispensaire, la pharmacie et la maternité.

Financement basé sur les résultats (FBR)

L'Agence de développement néerlandaise (SNV) et l'Institut royal des Tropiques (KIT) ont mis en place un projet pilote de financement basé sur les résultats (FBR) au Mali, plus précisément dans la région de Koulikoro, l'une des plus pauvres du pays (Ministère de la Santé du Mali 2012; Secrétariat général 2009; Toonen *et al.* 2014). Le projet a été implanté de février 2012 à décembre 2013 dans les districts sanitaires de Dioïla, Fana (tous deux dans le cercle de Dioïla) et Banamba (cercle de Banamba) et concernait un total de 26 CSCOM (18 pour le district de Dioïla et 4 pour chacun des districts de Fana et de Banamba) et 3 CSREF (1 pour chaque district) (Toonen *et al.* 2014).

Ce projet pilote avait pour but d'augmenter la qualité et la quantité de soins offerts dans ces districts sanitaires à travers des mécanismes de vérification, d'implication des communautés, de responsabilisation dans la gestion de fonds, de liaisons entre la motivation et le résultat, et d'achats (Toonen *et al.* 2014; Toonen, Kone et Dao s.d.). L'achat consistait en un paiement, sous forme de subventions, des services de santé ciblés par le projet qui évaluait ces services de manière qualitative et quantitative

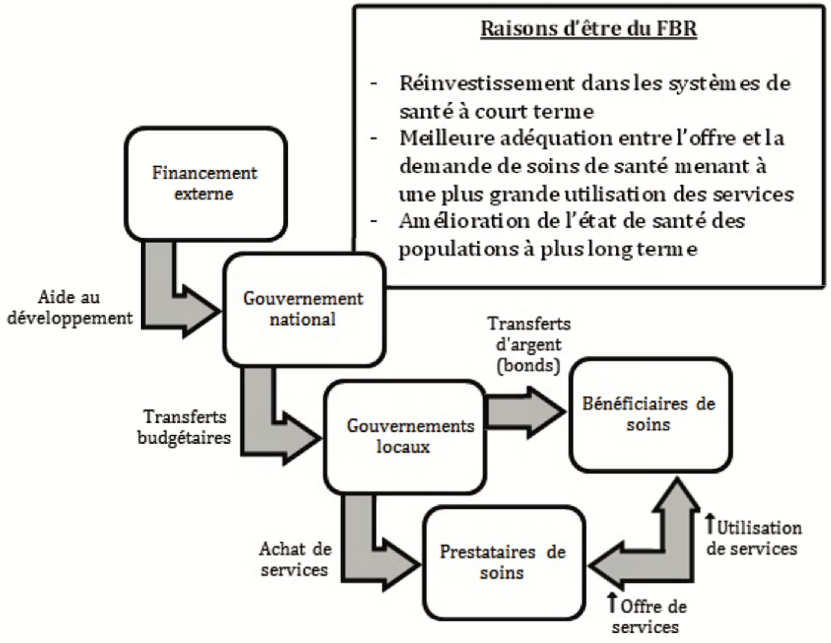


Figure 1. *Modèle logique du financement basé sur les résultats*

Ces subventions, communément appelées « primes », étaient distribuées aux centres de santé (60%) ainsi qu'aux prestataires de soins (40%). La proportion était inversée pour les CSREF. Ces primes étaient partagées entre le personnel de santé, chacun-e ayant obtenu en moyenne 140 500 FCFA par trimestre (pour un total de 65 419 921 FCFA/138 100 $CAN pour le projet), et les organisations de santé qui ont reçu un total de 98 129 882 FCFA (207 150 $CAN) (Toonen et al. 2014). Les primes individuelles constituaient une partie substantielle du revenu de la majorité des employé-e-s qui pouvaient ainsi jusqu'à doubler leur revenu (Toonen *et al.* 2014).

Divers acteurs et actrices étaient impliqués dans l'implantation du FBR, avec différents rôles : les équipes techniques des CSCOM et CSREF (prestations de soins), les mairies et les ASACO (mise en contrat et achat), l'équipe-cadre du district et la direction régionale de la santé (DRS) (régulation/vérification au niveau des CSCOM et des CSREF respectivement) et différentes organisations non gouvernementales (ONG) indépendantes (contre-vérification des résultats).

À la fin du projet (2014), une évaluation finale a été conduite sur la transmission des responsabilités au gouvernement malien et la mise à l'échelle (Toonen *et al.* 2014), mais sans aborder l'élément de la pérennité du projet. Il était donc pertinent d'évaluer la pérennité du projet FBR au Mali ainsi que le processus de pérennisation.

Plan d'évaluation

Dans cette étude, l'évaluation de la pérennité a porté sur deux objets : le processus de pérennisation et l'état actuel de la pérennité de l'intervention. En évaluant la pérennisation, on s'intéresse donc aux différents mécanismes permettant une pérennité future tandis qu'en évaluant la pérennité, on s'intéresse davantage aux acquis d'une intervention à un moment donné (Tableau 1).

Tableau 1 : Plan d'évaluation d'un projet FBR au Mali

	Pérennisation	**Pérennité**
Objectifs de l'évaluation	Comprendre le processus de pérennisation	Apprécier le niveau de pérennité du projet
Éléments de l'évaluation	Déterminants (pérennité) Phases (pérennité) Événements	Activités « routinisées » (via les déterminants et phases)
Collecte de données	Entretiens semi-structurés Recherche documentaire	
Type d'analyse	Appariement à un cadre conceptuel Élaboration d'une explication inductive	Appariement à un cadre conceptuel

Malgré cette différence, l'évaluation du processus et l'évaluation des acquis sont intimement liées puisqu'elles reposent toutes les deux sur une base commune; les déterminants de la pérennité sont la mémoire organisationnelle, les codes/valeurs, les règles/procédures, la capacité d'adaptation et les relations inter-acteurs et actrices (voir Seppey *et al.* 2017 pour plus d'informations sur ces déterminants). Une particularité de l'évaluation de la pérennité est qu'elle résulte d'un jugement de la part du chercheur ou de la chercheuse qui attribue un niveau de pérennité à l'intervention en lien avec les données collectées sur le terrain.

Méthodologie

Revue de la littérature et cadre conceptuel

L'évaluation d'un phénomène complexe nécessite préalablement une bonne compréhension de ses divers mécanismes, déterminants ou autres composantes, d'où la nécessité de commencer l'évaluation par une revue de la littérature de la pérennité. Pour cette étude de cas, plusieurs articles scientifiques, rapports d'organisation, présentations et autres documents ont contribué à une compréhension systématique et à l'élaboration d'un cadre conceptuel regroupant les divers éléments de la pérennité. Ces éléments sont présentés à la figure 2 : i) ses déterminants, ii) le processus de pérennisation, iii) le niveau de pérennité résultant des éléments précédents, iv) les différents contextes d'implantation de l'intervention, et v) des événements critiques et le temps (Johnson *et al.* 2004; Chambers, Glasgow et Stange 2013; Pluye, Potvin et Denis 2004; Moullin *et al.* 2015). La schématisation du cadre conceptuel permet donc de relier les différents sous-concepts de la pérennité (adaptation, contexte organisationnel, planification de la pérennité ou pérennité faible) à des mécanismes.

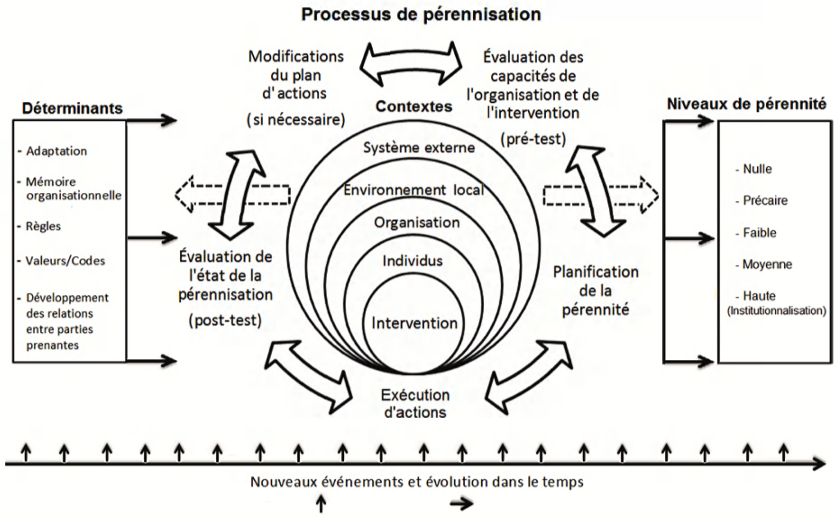


Figure 2. Cadre conceptuel de la pérennité. Adapté de Johnson et al. (2004), Chambers et al. (2013), Pluye, Potvin et Denis (2004) et Moullin et al. (2015).

Construction de l'outil et validation

L'outil de collecte de données est le guide d'entretien semi-structuré qui découle de la revue de littérature et du cadre conceptuel. Pour cette étude de cas, le modèle d'un guide préalablement utilisé lors d'une recherche sur la pérennité d'intervention en Haïti a servi de base (Ridde, Pluye et Queuille 2006). Pour mieux correspondre au cadre conceptuel, ce modèle a cependant dû être adapté avec l'ajout, le retrait et la modification de certaines questions.

Ce guide d'entretien compte cinq sections se rattachant aux différents déterminants du cadre et permet la collecte d'informations en lien avec différents événements critiques également liés au processus de pérennisation. Les questions étaient ouvertes et permettaient la collecte d'un maximum d'information en lien avec les thématiques abordées. Puisque le guide ayant servi de modèle avait été utilisé en Haïti, le nouveau guide a dû être adapté au contexte malien (Ridde, Pluye et Queuille 2006). Cette adaptation a été faite en deux temps, d'abord par l'entremise d'un collègue malien puis lors d'une visite sur le terrain où le guide a été testé avec de potentiels participants et participantes. Des modifications ont donc été faites quant au vocabulaire, à l'ordre des questions, à la complexité de certains concepts, etc. Cette étape d'adaptation du guide d'entretien est primordiale pour une collecte de données efficace.

Échantillonnage

Dans un but d'exhaustivité, une grande variété de centres de santé ont été sélectionnés comme sites de collecte, avec le soutien de personnes travaillant dans les CSREF. Plusieurs critères de sélection ont été utilisés pour garantir cette exhaustivité : des sites avec une haute ou basse fréquentation, en milieu urbain ou rural et avec des particularités dans l'implantation. Pour s'assurer de l'obtention de données riches et pertinentes, ont aussi été priorisés les sites avec un faible roulement de personnel, avec un accès physique sûr et avec un nombre plus élevé d'achats par FBR. Au total, l'échantillon des sites comprend trois CSCOM dans chacun des districts de Dioïla et de Fana sur les 26 possibles, ainsi que les deux CSREF respectifs.

Pour ce qui est des participant-e-s, deux types d'échantillonnage ont été utilisés. Le premier était raisonné et stratifié et mettait l'accent sur les différences entre les participant-e-s et l'exhaustivité des points de vue (Palinkas *et al.* 2015). Le critère d'inclusion à l'étude était d'avoir participé au projet et les strates étaient les différents rôles des participant-e-s : membres de la direction (9), prestataires de soins (22), membres de l'ASACO ou du conseil du cercle (9) et représentants locaux (4). Les directeurs des centres ont servi de personnes-ressources dans la sélection des participant-e-s en tant qu'autorités des différents centres, ceci facilitant l'accès aux participant-e-s (Savoie-Zajc 2007). L'autre type d'échantillonnage s'est fait par la méthode boule de neige qui consiste à obtenir l'aide d'un-e participant-e pour trouver un-e autre participant-e qui nous aidera ensuite à en trouver un-e autre, et ainsi de suite (Pires 1997). Ce type d'échantillonnage a permis d'obtenir plusieurs participant-e-s similaires : des personnes ayant participé à la conception du projet ainsi qu'à son implantation. Les participant-e-s provenaient de diverses organisations (Direction nationale et régionale de la santé, la SNV et le KIT; cinq au total) et étaient sélectionnés suivant la recommandation d'acteurs et actrices reliés au FBR préalablement rencontrés. L'échantillonnage final a permis d'atteindre un niveau de saturation des données avec un total de 49 entretiens. Trois éléments permettent d'affirmer la qualité de l'échantillonnage : 1) la représentativité en nombre de l'échantillon en lien avec la population à l'étude (par exemple, certains CSCOM disposaient de moins de 10 employés alors que le nombre cible de participant-e-s à l'étude était de cinq par CSCOM), 2) la participation de différents acteurs et actrices dans la sélection des participant-e-s (la coopération entre les chercheuses et chercheurs et divers acteurs et actrices terrains a pu diminuer les biais de sélection) et 3) l'exhaustivité des groupes d'acteurs et actrices (ex : médecin, sage-femme, gestionnaire, concepteur).

Collecte de données

L'étape de la collecte de données s'est déroulée en deux phases : du 22 au 29 décembre 2015, puis du 6 au 10 février 2016. Ces deux phases consistaient à se rendre (par bus, moto, charrette, auto-stop, etc.) aux différents sites qui pouvaient être très éloignés les uns des autres (plusieurs kilomètres de pistes à partir de la route goudronnée) et à rencontrer un maximum de

participant-e-s pendant une durée de un à trois jours. La deuxième phase de collecte était plus ciblée puisqu'elle consistait à rejoindre les différentes personnes n'ayant pas pu participer à la recherche lors de la première phase. Les différents contacts de ces personnes absentes ou ajoutées à l'échantillonnage par la méthode boule de neige ont été pris durant la première phase de collecte, ce qui a permis de faciliter la prise de rendez-vous et de diminuer le temps imparti à la collecte de données.

Exemple d'un dispensaire (Source: M. Seppey)

Tous les entretiens étaient enregistrés avec le consentement des participant-e-s et duraient de moins de 30 minutes à plus de 2 heures (± 1 h 15 minutes en moyenne). Ces entretiens se déroulaient majoritairement en français, mais plusieurs participant-e-s ont choisi de les faire en bambara, la langue locale. Un interprète accompagnait donc le chercheur et permettait de faire le pont entre le bambara et le français. Il est important de noter que plusieurs participant-e-s pouvaient faire leur entretien en français, mais avec des expressions ou des segments complets en bambara. La présence en tout temps de l'interprète était donc nécessaire.

Ces entretiens se déroulaient normalement au centre de santé qui était le milieu de travail des participant-e-s. Cependant, les entretiens pouvaient souvent être dérangés par des gens venant saluer l'équipe de recherche, des curieux, des vrombissements de motos, des chants de coq, etc., ce qui était typique du contexte local et du milieu de travail où il était généralement difficile de trouver un lieu tranquille pour l'entretien.

Exemple d'une difficulté d'accès à un centre de santé (Source: M. Seppey)

Une recherche documentaire a aussi été effectuée puisqu'il est

important que les données collectées via les entretiens soient corroborées par les documents officiels et *vice versa* ; les interventions étant souvent différentes sur papier que dans la réalité (Chambers, Glasgow et Stange 2013). La recherche documentaire s'est donc effectuée tout au long de l'étude. Une demande d'accès était faite pour chaque document mentionné dans les entretiens, mais très peu de documents ont été effectivement collectés à cause du faible niveau d'archivage dans les centres de santé et de la peur d'être mis en faute, entre autres.

Gestion, traitement et analyse des données

Mahamadou Diabaté (Assistant de recherche à MISELI) et Mathieu Seppey (Chercheur à l'IRSPUM) en camping aux centres de santé (Source: M. Seppey)

À la suite de la collecte de données, les versions audio des entretiens ont été transcrites. Les entretiens en bambara ont été transcrits par un agent de recherche tandis que ceux en français ont été transcrits par le chercheur principal. Pour vérifier l'intégrité de chaque transcription, le chercheur principal a dû relire tous les entretiens en français (30 entretiens sur les 49) avec l'enregistrement. Les enregistrements audio ainsi que les *verbatim* ont été enregistrés dans un système de stockage et de partage d'informations en ligne en format électronique. Les équipes de recherche au Mali et au Québec ont donc eu accès à ces informations.

L'intégralité des *verbatim* a ensuite été codée à l'aide du logiciel ©QDA Miner. Le codage avec ce logiciel consiste à attribuer des thématiques (ex. : adaptation, procédures, etc.) à différents segments provenant des *verbatim* (mots, phrases, paragraphes). Ce processus d'abstraction des données (passer de l'empirique à la théorie) est structuré par un arbre de codes qui est fondé sur le cadre conceptuel. L'arbre de codes a aussi été revu par un pair, ce qui a permis d'assurer la présence de chaque élément du cadre conceptuel.

En ce qui a trait à l'analyse, deux types ont été utilisés : l'analyse par appariement (l'association du cadre conceptuel avec les *verbatim*) et l'élaboration d'une explication via une analyse inductive et itérative des données permettant l'émergence de différents thèmes (Contandriopoulos *et al.* 1990; Yin 2013). L'analyse par appariement s'est faite sur la base des différents thèmes ayant été codés. Pour chaque composante du cadre conceptuel, on proposait des éléments empiriques illustrant la situation au Mali (ex. : l'incompréhension de certains termes médicaux illustre l'utilisation d'un langage inadapté). Les écarts entre le cadre conceptuel et les éléments empiriques ont aussi été notés ; ils étaient les principaux résultats de la recherche. Ensuite, une explication du phénomène de pérennité a été élaborée à travers une analyse itérative et inductive des données. Plusieurs éléments ont permis ce type d'analyse, dont la division en deux phases de la collecte de données : une première écoute et la transcription des enregistrements ont eu lieu avant le retour sur le terrain. Des hypothèses ont donc pu être infirmées ou confirmées durant la deuxième phase de collecte. L'écoute des enregistrements, la relecture des notes prises lors des entretiens, le codage des *verbatim*, la création de tableaux récapitulatifs par centre de santé et la présentation des résultats préliminaires sont des étapes qui ont permis de mieux s'approprier les données et de mieux analyser les différents aspects à l'étude. Une explication inductive a aussi pu être effectuée via les différentes thématiques (hors du cadre conceptuel) ayant été ajoutées à l'arbre de codes durant l'étape de codage.

Résumé des résultats

Les résultats de l'étude découlent des différentes étapes de la collecte de données. Pour plus de clarté, les résultats ont été confrontés au cadre conceptuel et aux thématiques ajoutées (Seppey *et al.* 2017). Nous avons conclu que le projet de FBR au Mali avait un niveau faible de pérennité puisque, selon notre évaluation, peu d'activités ont été « routinisées » au bout des deux ans suivant la fin de l'intervention, alors que c'était un des déterminants de la pérennité. Aussi, la continuité des activités FBR ayant perduré n'était pas assurée à court terme en raison de ressources instables, du manque d'implication de différents acteurs et actrices ou encore de l'absence de bénéfices clairs pour ces derniers. Concernant la pérennisation, plusieurs phases étaient insuffisantes ou absentes, par exemple la

planification et l'évaluation de la pérennité de l'intervention. Pour plusieurs acteurs et actrices, la pérennisation de l'intervention passait par sa mise à l'échelle, ce qui est cependant un tout autre processus. Des rumeurs à l'effet que la Banque mondiale prendrait en charge une mise à l'échelle du projet semblent avoir occulté la planification de la pérennisation du projet de FBR mis en place. Le processus de pérennisation a aussi été fortement influencé par l'absence de ressources stables dédiées au maintien de l'intervention. À la fin du financement du projet, plusieurs activités n'ont pas pu se poursuivre. Par exemple, sans ressources pour la supervision externe, plusieurs règles et procédures du FBR ont été abandonnées, ces dernières ayant été jugées trop exigeantes ou non nécessaires.

Analyse réflexive

Forces et faiblesses

La force de cette étude sur la pérennité est l'utilisation d'un cadre conceptuel qui a permis une compréhension plus systémique et exhaustive du phénomène. Dans le cas de l'évaluation d'un phénomène complexe, il nous semble indispensable d'utiliser un tel cadre pour ne pas occulter certaines dimensions pouvant affecter les résultats de l'étude. Ce cadre a aussi permis de mieux structurer toute la recherche, de la construction du guide d'entretien à la rédaction du rapport final. L'utilisation d'un cadre lors de l'analyse peut sembler très déductif, mais n'empêche toutefois pas l'utilisation d'une analyse inductive permettant à différents thèmes d'émerger (Blais et Martineau 2006).

Une limite de cette évaluation pourrait être l'absence d'autres méthodes de collecte de données, comme l'observation, pour valider certaines informations recueillies lors d'entretiens. Par exemple, il aurait été intéressant d'observer les différents soins offerts ainsi que les différentes pratiques des centres pour les confronter aux récits de services évoqués durant les entretiens. Plusieurs divergences auraient peut-être été soulevées entre ce qui était rapporté et ce qui était observé. Ceci peut être expliqué par le fait que les participant-e-s étaient majoritairement en situation de conflit d'intérêts puisqu'ils avaient profité directement ou indirectement du projet (à travers les primes du projet ou un emploi rémunéré en lien avec

le projet FBR). Cette situation s'est reflétée dans les données qui peuvent surestimer les effets et la pérennité du projet ou occulter les aspects négatifs ou non pérennes de ce dernier, bien que les résultats montrent globalement le contraire. En adoptant une méthode de collecte de données plus axée sur le discours (entretiens), les participant-e-s pouvaient avoir une plus grande propension à répondre favorablement quant aux questions en lien avec le projet. Des observations pendant et après le projet FBR auraient permis de voir effectivement les différents effets du projet FBR et sa pérennité.

Enjeux de la mise en œuvre de l'évaluation

Certains autres éléments peuvent constituer tant des limites que des appuis à l'évaluation. Le chercheur étant un *toubabou* (appellation courante d'un blanc) provenant de l'extérieur du contexte de l'étude et désirant discuter du thème du FBR, une proportion notable des participant-e-s le considérait comme un individu lié au FBR et pouvant réintroduire le projet. Cette préconception du rôle du chercheur peut avoir teinté négativement les données puisque la presque totalité des participant-e-s demandait à la fin de l'entretien le retour du projet. Malgré les stratégies mises en place pour diminuer ce biais, les participant-e-s étaient toujours sujets à de la désirabilité sociale (Savoie-Zajc 2016). Avant chaque entretien, le rôle du chercheur ainsi que les objectifs de l'étude étaient précisés et un accent était mis sur l'indépendance de l'étude face au projet FBR. Si les participant-e-s semblaient avoir des discours spécialement en faveur du FBR, un rappel était fait pour bien clarifier les objectifs de l'étude.

En contrepartie, le jeune âge du chercheur et son approche pourraient avoir facilité l'entrée sur le terrain. Étant plus jeune que la presque totalité des participant-e-s interviewés et utilisant des moyens de transport (auto-stop, motos, bus ou la marche) et des hébergements de base (tentes, salles d'hospitalisation, chambres de passage), l'équipe de recherche (l'interprète et le chercheur principal) s'est souvent attiré la sympathie de gens dans les centres de santé. Ceci a sûrement permis une meilleure disponibilité et disposition des participant-e-s ainsi qu'un certain niveau de confort avec l'équipe de recherche. L'utilisation du bambara par le chercheur principal lors de la prise de contact et de discussions informelles a aussi potentiellement permis d'établir un meilleur lien de confiance avec les participant-e-s.

Leçons apprises

Cette étude sur la pérennité permet de noter deux grandes leçons sur la manière d'effectuer un travail d'évaluation. En premier lieu, cette étude met en évidence l'importance de s'adapter au contexte, ce qui est valable tant pour la pérennité d'une intervention que pour le projet d'un chercheur ou d'une chercheuse! Par exemple, dans le cas de cette évaluation de pérennité, le langage utilisé pour les entretiens pouvait parfois être inadapté, malgré la validation préalable du guide sur le terrain. Les réponses très courtes et toujours positives de certains participant-e-s pourraient être expliquées par un mélange du désir de plaire et d'une incompréhension sûrement influencée par le faible niveau d'instruction de certains participant-e-s. Des formulations pouvaient aussi être mal adaptées. Par exemple, une « formation » dans le contexte malien est considérée comme une activité formelle se concluant généralement par un certificat ou un diplôme, ce qui n'était pas le cas pour le chercheur. Il devient donc primordial de clarifier les malentendus et incompréhensions potentielles. Ceci peut être facilité par l'intervention d'un-e interprète ayant été bien formé-e sur le sujet à l'étude ou par un choix de vocabulaire adapté pour chaque participant-e. Dès qu'un doute se crée quant à une potentielle confusion, le chercheur ou la chercheuse doit clarifier la situation et prendre note de cette nouvelle information lors des entretiens subséquents. L'outil de collecte doit donc être constamment adapté, tant le guide d'entretien que l'équipe de recherche elle-même à travers son vocabulaire ou ses comportements.

D'un point de vue plus général, la logistique d'une évaluation en santé mondiale demande souvent beaucoup de flexibilité et une grande capacité d'adaptation à cause de circonstances imprévisibles (transport en panne, maladie, perte de réseaux de communication, etc.). Il est toutefois possible de les mitiger en effectuant une bonne planification du travail à effectuer et du temps à y impartir. Planifier une journée pour effectuer 90 km de transport peut sembler ridicule, mais en ajoutant la panne d'autobus, les arrêts fréquents, les contrôles routiers, etc., la journée devient vite remplie. Planifier trop de rencontres en une journée peut aussi poser problème puisque des absences, des reports ou des invitations à divers événements non planifiés peuvent survenir. Prendre et planifier du temps « libre » est donc essentiel pour tirer le meilleur des différentes opportunités ou inconvénients. Diviser son étude en différents moments peut aussi aider

à recueillir plus d'informations (ex : planifier une deuxième plage d'horaire pour rencontrer les participant-e-s absents ou ajoutés à l'échantillon) et débuter plus tôt le processus d'analyse. Une première phase d'analyse peut ainsi permettre au chercheur ou à la chercheuse de retourner voir divers acteurs et actrices et d'obtenir des réponses aux premières questions. Le processus itératif de l'analyse permet de recueillir de l'information plus riche et de valider les nombreuses hypothèses préliminaires.

Références clés

Site http://www.etudiantsetdeveloppement.org/article/comment-assurer-la-perennite-de-votre-projet-de-si

Ce site a été créé par un réseau français Étudiants et Développement qui met de l'avant le transfert de connaissances en lien avec le développement entre pair-e-s. En plus de la pérennité, plusieurs thématiques en lien avec le développement sont présentées.

Site https://evalorix.com/boutique/innovations-en-sante/evaluation-perennite-programmes-sante-publique/

Evalorix.com est un site où se retrouve une multitude d'outils servant à l'amélioration des pratiques en santé et autres domaines. Ici, on peut trouver des outils d'évaluation concernant la pérennité.

Ridde, V., Pluye, P. et Queuille, L. (2006). Évaluer la pérennité des programmes de santé publique : un outil et son application en Haïti. *Revue d'Épidémiologie et de Santé Publique*, 54, 421-431.

Cet article est un exemple d'étude de cas portant sur l'évaluation de la pérennité d'une intervention en santé. Le modèle de guide d'entretien fourni à la fin de cet article a servi pour l'étude de cas effectuée au Mali.

Blais, M. et Martineau, S. (2006). L'analyse inductive générale : description d'une démarche visant à donner un sens à des données brutes. *Recherches Qualitatives*, 26(2), 1-18.

Cet article présente une introduction très concrète de ce qu'est l'analyse qualitative (et un peu plus spécifiquement l'analyse inductive). L'article clarifie les différentes étapes de l'analyse inductive tout en présentant un résumé d'autres types d'analyses.

Brouselle, A., Champagne, F., Contandriopoulos, A.-P. et Hartz, Z. (2011). *L'évaluation : concepts et méthodes*. Montréal : Les Presses de l'Université de Montréal.

Ce livre présente une diversité d'évaluations pouvant être utilisées dans différents contextes. Chaque type d'évaluation est donc explicité avec ses objectifs, les éléments évalués, le processus d'évaluation, etc.

Références

Audibert, M. et de Roodenbeke, E. (2005). *Utilisation des services de santé de premier niveau au Mali : analyse de la situation et perspectives*. Banque Mondiale, Région Afrique, Département du Développement Humain.

Blais, M. et Martineau, S. (2006). L'analyse inductive générale : description d'une démarche visant à donner un sens à des données brutes. *Recherches Qualitatives*, 26(2), 1-18.

Chambers, D., Glasgow, R. et Stange, K. (2013). The dynamic sustainability framework: addressing the paradox of sustainment amid ongoing change. *Implementation Science*, 8 (1), 117.

Contandriopoulos, A.-P., Champagne, F., Potvin, L., Denis, J.-L. et Boyle, P. (1990). *Savoir préparer une recherche : la définir, la structurer, la financer*. Montréal: Presses de l'Université de Montréal.

Johnson, K., Hays, C., Center, H. et Daley, C. (2004). Building capacity and sustainable prevention innovations: a sustainability planning model. *Evaluation and Program Planning*, 27(2), 135-149. doi: http://dx.doi.org/10.1016/j.evalprogplan.2004.01.002.

Lamoureux, A. (2003). *Recherche et Méthodologie en Sciences Humaines* (2e éd.). Laval: Éditions Études Vivantes.

Ministère de la Santé du Mali. (2012). *Carte sanitaire du Mali mise à jour – 2011 : Rapport de synthèse*. edited by Secrétariat Général. s.l.

Moullin, J., Sabater-Hernandez, D., Fernandez-Llimos, F. et Benrimoj, S. (2015). A systematic review of implementation frameworks of innovations in healthcare and resulting generic implementation framework. *Health Research Policy and Systems*, 13 (1), 16.

Observatoire africain de la santé. (2014). État de santé et tendances. Repéré

à http://www.aho.afro.who.int/profiles_information/index.php/Mali : Health_Status_and_Trends/fr.

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N. et Hoagwood, K. (2015). Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(5), 533-544. doi: 10.1007/s10488-013-0528-y.

Pires, A. (1997). Échantillonnage et recherche qualitative : essai théorique et méthodologique. *La recherche qualitative. Enjeux épistémologiques et méthodologiques*, 113-169. Montréal: Gaëtan Morin.

Pluye, P., Potvin, L. et Denis, J.-L. (2004). Making public health programs last: conceptualizing sustainability. *Evaluation and Program Planning*, 27(2), 121-133.
http://dx.doi.org/10.1016/j.evalprogplan.2004.01.001

Ridde, V., Pluye, P. et Queuille, L. (2006). Évaluer la pérennité des programmes de santé publique : un outil et son application en Haïti. *Revue d'Épidémiologie et de Santé Publique*, 54(5), 421-431.
http://dx.doi.org/10.1016/S0398-7620(06)76740-2

Savoie-Zajc, L. (2007). Comment peut-on construire un échantillonnage scientifiquement valide?. *Recherches qualitatives*, 99-111(5).

Savoie-Zajc, L. (2016). L'entrevue semi-dirigée. Dans B. Gauthier et I. Bourgeois (dir.), *Recherche sociale : de la problématique à la collecte des données* (6e éd., p. 337-364). Québec : Presses de l'Université du Québec.

Secrétariat général du Ministère de la Santé du Mali. (2009). *PRODESS II prolonge 2009 2011* : Composante santé.

Seppey, M., Ridde, V., Touré, L. et Coulibaly, A. (2017). Donor-funded project's sustainability assessment: a qualitative case study of a results-based financing pilot in Koulikoro region, Mali. *Globalization and health*, 13(1), 86.

Syll, O. (2005). *La décentralisation en Afrique subsaharienne : Mali* (Mémoire de maîtrise). Besançon: Université de Franche-Comté.

Toonen, J., Dao, D., Matthijssen, J. et Koné, B. (2014). *Évaluation finale : accélérer l'atteinte de l'OMD 5 dans la région de Koulikoro – Projet pilote financement basé sur les résultats dans les cercles de Dioïla et Banamba*. Amsterdam: Institut Royal Tropical.

Toonen, J., Koné, B. et Dao, D. s.d. Le Financement Basé sur les Résultats (FBR) au Mali. KIT.

United Nations Development Programme. (2015). *Human Development Reports.* http://hdr.undp.org/en/countries/profiles/MLI

United Nations Statistics Division. (2016). *Country Profile: Mali.* http://data.un.org/CountryProfile.aspx?crName=Mali.

Yin, R.K. (2013). *Case Study Research: Design and Methods*. Thousand Oaks: Sage Publications.

Résumé / Abstract / Résumé en bambara

Le financement basé sur les résultats (FBR) est un type de réforme du financement du système de santé très populaire en Afrique. Plusieurs projets pilotes sont mis en œuvre dans le but d'offrir plus de soins de santé ainsi que d'en améliorer la qualité. Cependant, la pérennisation de ce type d'intervention en Afrique a encore été trop peu étudiée. Ce chapitre explique donc comment la pérennité de cette intervention a été évaluée au moyen d'une méthodologie qualitative. En lisant ce chapitre, on pourra mieux comprendre le processus de recherche qui a été mis en œuvre grâce à sa présentation exhaustive, de la première étape (ex. : la réalisation d'une revue de la littérature) jusqu'à la dernière (ex. : l'attribution du niveau de pérennité de l'intervention). Une section plus réflexive permet aussi aux lecteurs et lectrices de s'immiscer davantage dans l'expérience des chercheurs en ayant une meilleure compréhension des défis rencontrés (tant méthodologiques que professionnels) et des leçons apprises.

***

Results-Based Financing (RBF) is a new type of health financing reform that is very popular in Africa. Several pilot projects are being implemented on this continent, and elsewhere, to provide better and more health care services. However, too little research has been done on the sustainability of this type of intervention in Africa. This chapter explains how this intervention was evaluated, through the concept of sustainability, and following a qualitative methodology. By reading this chapter, we will be able to better understand the research process that has been carried out

thanks to its exhaustive presentation, from the first step (e.g. the completion of a review of the literature) to the last (e.g. the attribution of the level of sustainability of the intervention). A more reflexive section also allows readers to become more involved in the researcher's experience by having a better understanding of the challenges encountered (both methodological and professional) and the lessons that they have learned.

***

Résumé en bambara, traduit par Ahamadou Diakité et Mahamadou Diabaté que nous remercions

Nafolo ko mi siysin né do diabi sɔrɔlila (FBR) yé nafolo taabolo kura yé mi maakadi keneya kan farafina kɔnɔ. Poroje kuntiléna ciaman be sékan walassa ka sé ko furakeli lagini boŋan ka yɛlɛ mali do a la ani ka kɛcioŋoya ŋuma kɛ o hokumu kɔnɔ taabolo bilali sé kan kuolaye farafina kɔnɔ kalan ciamaba ma a siŋi sé kan o la. Ni Bolofara ŋinfoli ba yira ciokodi taabolo kudayi diatemilɛli kɛra ni baara minɛ, ŋinini taabolo bɔlé yuman na. Ni Bolofara kalan ni na, ŋinini taabolo fouamu bibali la sé kan sannan yɛlin yirali la ni kalan biolo folonŋi (Misali la : o ŋafé sebeni waleyali la) fo lavan (Misali la : taabolokudayi diaté dero dili la). Hakilina baara bolodali boyan k ata a fɛ ki sé do a ko la kɔsɔbɛ ŋinini kela yé dɔniya mi sɔrɔ ŋeleya minounou yéra faamu cioko yuman na (kaata o kalan taabolo ni o baara taabolo) ani fouamu mi sɔrɔ la kɔnɔ.

***

Mathieu Seppey a découvert le monde de la recherche après avoir passé quelque temps à faire de l'aide humanitaire. C'est en voyant différents problèmes sur le « terrain » qu'il a constaté la nécessité de la recherche et l'importance de mieux comprendre les différents mécanismes que l'on actionne en effectuant divers projets ou programmes d'intervention. Mathieu est donc retourné sur les bancs d'école pour mieux comprendre les interventions mises en place partout dans le monde, et plus spécifiquement en Afrique de l'Ouest. Effectuant une maîtrise en administration de la santé, il s'est rendu au Mali dans le but de mieux comprendre ces mécanismes et plus spécialement ceux de la pérennisation d'un projet de financement basé sur les résultats (le sujet de son mémoire). Aujourd'hui, il poursuit ses recherches au doctorat, sa thèse étant sur la définition de la pérennité d'interventions en santé.

Valéry Ridde est directeur de recherche au CEPED[1], une Unité Mixte de Recherche associant l'Université Paris Descartes et l'Institut de Recherche pour le Développement (IRD). Il est chercheur régulier de l'Institut de recherche en santé publique de l'Université de Montréal (IRSPUM) et chercheur associé au Centre interdisciplinaire de recherche en développement international et société (CIRDIS) de Montréal. Il est rédacteur adjoint des revues *BMJ Global Health* et *Global Health Promotion*, membre du comité de rédaction de la *Revue Canadienne d'Évaluation de Programme* et de *Qualitatif! Revue africaine de sciences sociales*. Ses travaux de recherche portent sur la couverture universelle en santé, le financement des services de santé, l'évaluation de programme, les politiques publiques de santé et la promotion de la santé.

Citation

Mathieu Seppey et Valéry Ridde. (2019). L'évaluation de la pérennité. Une intervention de financement basé sur les résultats au Mali. In *Évaluation des interventions de santé mondiale. Méthodes avancées*. Sous la direction de Valéry Ridde et Christian Dagenais, pp. 33-53. Québec : Éditions science et bien commun et Marseille : IRD Éditions.

1. http://www.ceped.org

PARTIE II

LES APPROCHES QUALITATIVES ET PARTICIPATIVES

3. L'évaluation qualitative, informatisée, participative et inter-organisationnelle (EQUIPO)

Exemple d'un programme en faveur des femmes victimes de violences en Bolivie

MATHIEU BUJOLD ET JEAN-ALEXANDRE FORTIN

La méthode ÉQUIPO

Définition de la méthode

Elle comporte trois volets (qualitatif, participatif et inter-organisationnel) et suit six étapes principales :

1. Introduction au terrain et constitution des équipes d'évaluation
2. Adaptation du matériel pédagogique au contexte socioculturel
3. Formation d'introduction aux méthodologies qualitatives
4. Recrutement des participant-e-s et collecte de données qualitatives
5. Analyse qualitative assistée par ordinateur
6. Forum inter-organisationnel : un espace de partage

Forces de la méthode

- Plusieurs niveaux de répercussion potentielle : institutionnel, personnel, technique, professionnel, collaboratif
- Le volet qualitatif favorise l'apprentissage des forces et des faiblesses des programmes et des ONG du point de vue des bénéficiaires et permet d'ajouter une dimension humaine aux indicateurs quantitatifs
- Le volet participatif motive les employé-e-s et les bénéficiaires en renforçant leur autonomisation tout en favorisant leur participation démocratique au processus d'évaluation d'un programme auquel ils et elles participent
- Le volet inter-programmatique et inter-organisationnel offre une perspective globale sur les façons d'améliorer les programmes

Défis de la méthode

- Mobilisation spontanée des employé-e-s (non imposée par la direction)
- Planification de ressources humaines pour l'évaluation et les expériences des organisations
- Participation des bénéficiaires à toutes les étapes de l'évaluation
- Appui d'une organisation qui applique elle-même la méthode ÉQUIPO
- Convaincre les bailleurs de fonds de la plus-value des méthodes qualitatives

Un programme est un ensemble organisé et cohérent d'activités dont l'objectif est de produire des changements pour les participant-e-s potentiels (Plante 1994; Ridde et Dagenais 2012b). Les programmes de développement visent spécifiquement à produire des changements dans l'existence de personnes touchées par des problématiques souvent reliées à des facteurs socioculturels complexes (par exemple les relations inter-genres). Si les modèles logiques sont des outils essentiels à la planification et à l'évaluation de programmes (Porteous 2012), les réalités du terrain sont souvent beaucoup plus complexes que prévu. Les processus de changement ne peuvent être résumés à une logique linéaire de cause à effet et leur évaluation nécessite la participation d'une multitude de parties prenantes issues de différentes organisations. Les trois volets de la méthode ÉQUIPO tiennent compte de ces réalités complexes.

Le volet qualitatif d'ÉQUIPO

Le volet qualitatif de cette méthode vise à comprendre les processus de changement ou d'inertie depuis le point de vue de celui ou celle qui expérimente une situation donnée (par exemple la participation, en tant qu'employé-e ou bénéficiaire, à un programme visant à réduire la violence). Selon les parties prenantes (bénéficiaires, technicien-e-s, direction), comment et pourquoi un programme a mené ou non à des changements dans l'existence des bénéficiaires ? Quelles sont les forces et les faiblesses du programme ? Comment peut-il être amélioré dès maintenant ? Quels sont les défis à surmonter ? Des méthodes qualitatives peuvent être mobilisées pour explorer ces questions.

La collecte de données qualitatives, par exemple par l'entremise d'entrevues semi-dirigées ou de groupes de discussion, est de plus en plus utilisée en évaluation de programmes. Cependant, l'analyse de ce type de données est un défi de taille (Patton 2015). L'analyse thématique est une méthode accessible pour tirer des significations d'un corpus de données qualitatives (Miles, Huberman et Saldaña 2014; Paillé 2016). Pour décrire simplement cette méthode, faisons un parallèle avec une lecture systématique de textes où les passages importants sont surlignés puis marqués dans la marge par des codes thématiques pour se rappeler de leurs contenus. De façon intuitive, l'évaluatrice ou l'évaluateur attentif fait de l'analyse thématique lorsqu'il ou elle lit des rapports ou des verbatims

d'entrevues, surligne des extraits pertinents et prend note de leurs contenus. Il rassemble ensuite ces informations par groupes dans un document présentant le résultat de son analyse. Cependant, la tâche se complique lorsqu'il faut faire une analyse systématique de plusieurs documents et suivre une multitude de thèmes (par exemple les accomplissements, les défis, les facilitateurs, les recommandations). L'analyse thématique assistée par ordinateur est un processus rigoureux d'organisation de ces extraits de textes surlignés qui sont découpés et regroupés dans des catégories thématiques (par exemple les défis) pouvant contenir des sous-catégories (par exemple les défis technologiques). Dans un processus déductif, ces catégories sont prédéterminées, alors que, dans un processus inductif, elles émergent au fil de l'analyse.

Les critères de qualité de la recherche ou de l'évaluation qualitative diffèrent de ceux des approches quantitatives (Laperrière 1997; Patton 1999). Si elle n'est pas garante d'une analyse de qualité, l'utilisation de CAQDAS (*Computer-Assisted Qualitative Data Analysis Software*), comme le logiciel NVivo, peut augmenter la rigueur de l'analyse thématique (Bazeley et Jackson 2013; Bujold 2016). L'apprentissage de ce type de logiciels fait partie des programmes de formation continue offerts aux évaluateurs et évaluatrices de programmes.

Le volet participatif d'ÉQUIPO

Ce volet rappelle que l'évaluation de programmes n'est pas une fin en soi, mais un moyen d'améliorer les conditions sociales et collectives. L'évaluation participative cherche à renforcer l'autonomisation des parties prenantes en les faisant participer activement au processus d'évaluation qui se fait entre elles plutôt que sur elles (Cargo et Mercer 2008; Jagosh *et al.* 2012). Si le volet qualitatif permet d'explorer les points de vue et les expériences vécues des bénéficiaires et des technicien-ne-s des programmes, le volet participatif d'ÉQUIPO outille les personnes les plus motivées pour former des équipes de co-évaluateurs et co-évaluatrices afin de mener les différentes étapes du projet d'évaluation (planification, collecte et analyse de données, diffusion des résultats). Cette approche vise également à changer la perception négative envers l'évaluation de programmes qui est souvent réduite à la reddition de comptes et perçue comme un contrôle de performance de la part des responsables d'un programme (Ridde et Dagenais 2012a). Le niveau

de participation des co-évaluateurs et co-évaluatrices peut aller de la simple consultation à leur implication à toutes les étapes de l'évaluation (Daigneault et Jacob 2009).

Le volet inter-organisationnel d'ÉQUIPO

Ce volet élargit l'aspect participatif en stimulant les échanges de méthodes et de résultats d'évaluation de programmes entre des ONG qui partagent des missions similaires. ÉQUIPO vise le travail d'équipe entre parties prenantes d'organisations pour mieux affronter des problématiques socioculturelles complexes, comme la justice entre les femmes et les hommes.

Encadré 1. Les prérequis à l'application d'ÉQUIPO

- Répondre à un besoin en évaluation qualitative exprimé par les ONG locales.
- Obtenir le soutien de la direction des ONG dans le but de favoriser la participation des employé-e-s et des bénéficiaires.
- Pouvoir offrir des formations et un accompagnement d'un conseiller ou d'une conseillère en analyse qualitative et participative.
- Faciliter l'utilisation d'outils informatiques par les participant-e-s.
- Compter sur la disponibilité de ressources des ONG (humaines, matérielles et temporelles).
- Obtenir un engagement de collaboration inter-organisationnelle des ONG participantes.

Contexte : Justice entre les femmes et les hommes en Bolivie

Le projet pilote ÉQUIPO a été réalisé dans le cadre du programme de coopération volontaire « Mieux agir et mieux influencer » exécuté par Oxfam-Québec entre 2009 et 2015 dans 11 pays, dont la Bolivie. L'un des secteurs d'intervention du programme était celui de la « Justice entre les femmes et les hommes ». L'exercice de la violence, le féminicide et le harcèlement des femmes au travail sont encore fréquents en Bolivie (Ministerio de justicia 2015). En raison de l'importance de cette problématique, Oxfam-Québec s'est associé avec le *Centro Juana Azurduy* (CJA) et le *Centro para las mujeres Gregoria Apaza* (CPMGA), deux ONG locales ayant des programmes de promotion de l'égalité entre les femmes et les hommes.

Tableau 1. Descriptif des ONG partenaires

.	***Centro de Promoción de la Mujer Gregoria Apaza***	***Centro Juana Azurduy***
Mission	Contribuer à la reconnaissance et à l'exercice des droits des femmes.	Contribuer à l'autonomisation des femmes, individuellement et collectivement, en renforçant leurs capacités de représentation, d'exigibilité de leurs droits et de la redistribution équitable des ressources disponibles.
Localisation (données 2012)	El Alto, Bolivie. Altitude 4150m. Population : 848 452	Sucre, Bolivie. Altitude 2780m. Population : 261 201
Fondation	8 juin 1984	10 septembre 1989
Début partenariat avec Oxfam-Qc	2005	2003
Nombre d'employés (2012)	40 (29F/11H)	50 (30F/20H)
Nombre de bénéficiaires (estimation 2012)	2 000	1 500
Populations cibles	Urbaine. Femmes et jeunes. Métisses et Aymara.	Urbaine et rurale. Femmes et jeunes. Métisses et Quechuas.

C'est à la demande du CJA et du CPMGA qu'Oxfam-Québec recruta en février 2012 un conseiller en analyse qualitative (Mathieu Bujold) pour combler un besoin de formation en méthodes qualitatives et participatives d'évaluation. Un autre besoin exprimé par les partenaires était de stimuler la collaboration avec d'autres organisations boliviennes aux missions similaires. C'est dans cet objectif que les membres du personnel du CJA et du CPMGA se sont réunis en mars 2012. Ces derniers ont exprimé la nécessité d'être outillés pour mener conjointement un projet d'évaluation qualitative et participative de programmes d'attention aux femmes victimes de violence. Le projet pilote d'Évaluation QUalitative Informatisée et Participative inter-organisationnelle (ÉQUIPO) a été proposé aux partenaires pour répondre à ce besoin.

Tableau 2. Les objectifs d'ÉQUIPO

Objectif général	Renforcer les capacités des partenaires (ici, CJA et CPMGA) en méthodes qualitatives et participatives d'évaluation.
Sous-objectifs qualitatifs	Former les parties prenantes (bénéficiaires, technicien-ne-s et direction) des ONG partenaires à la collecte et à l'analyse de données qualitatives assistée par ordinateur (NVivo). Générer une base de données qualitatives pour explorer les expériences vécues par les bénéficiaires des ONG partenaires, recueillir leur point de vue sur les programmes, ainsi que leurs recommandations. Favoriser l'apprentissage dans l'action, en évaluant les processus des programmes des ONG partenaires.
Sous-objectifs participatifs	Développer les compétences des parties prenantes en évaluation participative de programme. Favoriser la participation active et démocratique des parties prenantes, à titre de co-évaluateurs et co-évaluatrices de programme des organisations partenaires.
Objectifs inter-organisationnels	Promouvoir des échanges entre des ONG ayant des missions similaires ou complémentaires. Partager les expériences vécues et les défis rencontrés lors de l'application de la méthode ÉQUIPO. Renforcer une culture participative d'évaluation et de suivi des ONG partenaires.

Budget et engagement des partenaires

- Appui d'Oxfam-Québec
- Un conseiller, Mathieu Bujold (MB), en analyse de données qualitatives.
- Un conseiller, Jean-Alexandre Fortin (JAF), en développement organisationnel.
- Deux licences complètes du logiciel NVivo.
- Deux manuels de formation traduits en l'espagnol.
- Soutien financier couvrant 75 % des coûts de la collecte de données qualitative, de la transcription et de la traduction en espagnol des groupes de discussion et des entrevues semi-dirigées réalisées en quechua ou en aymara.
- Implication des ONG locales

- Recrutement des co-évaluateurs et co-évaluatrices (employé-e-s, stagiaires et bénéficiaires).
- Apport financier couvrant 25 % des coûts relatifs à la transcription et à la traduction.
- Enregistreuses audio et ordinateurs portables pour l'installation du logiciel NVivo.
- Frais de déplacement et une collation offerte aux participant-e-s lors des groupes de discussion et des entrevues.

ÉQUIPO : une méthode d'apprentissage dans l'action

Introduction au terrain et constitution des équipes d'évaluation

À son arrivée, Mathieu Bujold (MB), l'expert en analyse de données qualitative, a été jumelé à Jean-Alexandre Fortin (JAF), expert terrain et conseiller en développement organisationnel d'Oxfam. JAF travaillait depuis plus de quatre ans avec les deux ONG locales. Pour commencer, des rencontres ont été planifiées avec les directions des ONG et les employé-e-s intéressé-e-s, afin de mieux saisir leurs besoins et leurs attentes envers les approches qualitatives et participatives. Chaque ONG a choisi trois programmes à évaluer (CJA : *Defensoría de las mujeres*, *Acción Política*, *Trabajo Digno* ; CPMGA : *Servicio psico Legal*, *Brigadistas*, *Agentes Comunitarias*). Des équipes d'évaluateurs et d'évaluatrices ont ensuite été constituées pour participer au projet pilote ÉQUIPO. La consigne générale était qu'au moins un-e employé-e des programmes concernés participe à l'évaluation d'un programme autre que le sien. Chaque équipe d'évaluateurs et d'évaluatrices a ensuite réfléchi à des bénéficiaires potentiellement intéressé-e-s à participer au processus d'évaluation.

Adaptation du matériel pédagogique

Le conseiller en analyse qualitative a partagé des manuels de formation (Bujold 2014 et 2016) et plusieurs guides méthodologiques qui ont été traduits en espagnol (réalisation de groupes de discussion et d'entrevues ; guide de transcription ; protocole de validation inter-analystes). Puisque ces outils avaient été développés pour la recherche universitaire, ils ont dû

être adaptés au contexte d'évaluation de programmes de développement. Par exemple, des aspects théoriques ont été simplifiés, et l'apprentissage dans l'action a été favorisé. Malgré ces assouplissements méthodologiques, le respect des critères de qualité fondamentaux de l'évaluation qualitative a été maintenu, par exemple l'optimisation de l'adéquation empirique entre les réalités évaluées, les données collectées et les productions d'évaluation (Olivier de Sardan 2008). Un processus de validation inter-analystes mobilisant un groupe interdisciplinaire d'analystes pour interpréter un corpus de données a aussi été appliqué (Miles, Huberman et Saldaña 2014).

Formation d'introduction aux méthodologies qualitatives

Puisque l'utilisation de CAQDAS nécessite une compréhension des méthodologies qualitatives, deux jours de formation sur le sujet ont d'abord été offerts aux employé-e-s et aux stagiaires des ONG. L'enseignement du logiciel NVivo s'est fait lors des étapes de collecte et d'analyse de données. Des formations spécifiques ont été offertes aux facilitateurs et observateurs, hommes et femmes, des groupes de discussion (co-évaluateurs et co-évaluatrices employé-e-s). Le rôle des facilitateurs et des facilitatrices est de stimuler la discussion pour explorer en profondeur les thèmes de discussion et favoriser la participation équitable des femmes et des hommes. Les observateurs et les observatrices ont la responsabilité de noter les aspects inaudibles à l'audio (par exemple les émotions) et surtout la chronologie des échanges (qui parlait, et quand). Les notes de l'observateur ou de l'observatrice sont essentielles à la personne responsable de la transcription. Des formations spécifiques ont également été offertes aux personnes chargées des interviews (co-évaluateurs et co-évaluatrices bénéficiaires) ainsi que des transcriptions (co-évaluateurs et co-évaluatrices stagiaires).

Recrutement des participant-e-s et collecte de données qualitatives

Pour chaque programme à évaluer, deux groupes de discussion ont été planifiés : le premier avec une cohorte de nouvelles bénéficiaires et le deuxième avec des femmes ayant récemment complété le programme (< 6 mois). Compte tenu de la charge émotionnelle reliée à la thématique, les groupes de discussion ont été menés par les employé-e-s des ONG

plutôt que par les bénéficiaires. Un ou une psychologue était présent-e dans chaque groupe. Quatre bénéficiaires ont été formé-e-s puis accompagné-e-s dans la planification et la réalisation des entrevues individuelles semi-structurées avec d'autres bénéficiaires. Les guides de discussion et d'entrevues semi-dirigées ont été élaborés conjointement entre les employé-e-s et les bénéficiaires. Si certains groupes de discussion avaient des thèmes de discussion spécifiques, par exemple la perception de la violence, tous les groupes devaient aborder des thèmes génériques (encadré 2).

Encadré 2. Thèmes de discussion génériques aux groupes de discussion d'ÉQUIPO

1. Pourquoi êtes-vous venue au centre initialement ?
2. Est-ce que votre expérience avec le centre a changé quelque chose dans votre vie ?
3. Êtes-vous satisfaite des résultats des formations reçues ?
4. Selon vous, quels sont les points forts et les faiblesses du programme ?
5. Comment le centre pourrait-il améliorer ses services ?

Les groupes de discussion étaient d'une durée d'environ deux heures, alors que les entrevues duraient en moyenne 30 minutes. Chaque activité de collecte de données a été enregistrée, puis transcrite en version électronique afin de faciliter l'analyse thématique avec NVivo. Compte tenu de la sensibilité du sujet, l'enregistrement audio a été préféré au format vidéo, ce qui rendait le rôle de l'observateur ou de l'observatrice d'autant plus important pour noter qui parlait, et à quel moment. La confidentialité étant particulièrement importante dans le contexte (violence), il a été expliqué aux participant-e-s que leur nom serait modifié et qu'il serait impossible de reconnaître leurs propos dans les rapports finaux. Les aspects éthiques de la recherche qualitative ont fait l'objet d'une formation auprès des co-évaluateurs et co-évaluatrices. Des groupes de discussion dans les communautés éloignées ont été réalisés en quechua ou en aymara et ont nécessité l'emploi de personnes chargées de la traduction et de la transcription. Un total de 11 groupes de discussion et de 14 entrevues rejoignant plus de 80 bénéficiaires de six programmes ont été réalisés, enregistrés, transcrits et analysés par les équipes d'analystes constituées d'employé-e-s du CJA et du CPMGA.

Analyse qualitative assistée par ordinateur

Suivant une démarche d'analyse thématique hybride, à la fois déductive et inductive (Fereday et Muir-Cochrane 2006), la codification assistée par NVivo des verbatims a permis de faire émerger des catégories thématiques du discours des bénéficiaires tout en portant attention aux indicateurs prévus par les organisations. Un processus rigoureux de validation inter-analystes de la codification a été appliqué. Chaque groupe a d'abord élaboré un dictionnaire de codification définissant les catégories et sous-catégories thématiques envisagées et donnant un exemple d'un extrait fictif ou réel qui y serait codifié. Les analystes ont appliqué cet outil en analysant individuellement un même groupe de discussion avec NVivo. Les membres de l'équipe se sont ensuite réunies avec le conseiller en analyse pour comparer leur codification et optimiser le dictionnaire de codification en ajoutant, fusionnant ou supprimant des catégories et/ou des sous-catégories. Après l'atteinte d'une certaine homologie de codification, c'est-à-dire quand les analystes codifiaient globalement les mêmes extraits de textes dans les mêmes catégories, le reste du corpus de données a été distribué entre les analystes. Ces derniers se sont ensuite rencontrés hebdomadairement en compagnie du conseiller pour présenter leurs analyses et mettre à jour le dictionnaire de codification. Après avoir codifié le corpus de données entièrement, une étape d'interprétation a été enclenchée. Les analystes se sont divisé la tâche de relire les extraits codifiés dans chacun des thèmes et de noter, dans des mémos, leurs interprétations en les illustrant de citations des bénéficiaires pour ainsi optimiser l'adéquation empirique entre les réalités explorées, les données générées et les produits d'évaluation. Pour chaque catégorie, des graphiques représentant la distribution des codifications par nombre de participant-e-s étaient ajoutés aux mémos. Ces mémos ont servi à l'élaboration de présentations pour le forum inter-organisationnel.

Forum inter-organisationnel : un espace de partage

En octobre 2012, les équipes d'analystes se sont réunies à El Alto, dans les bureaux du CPMGA, pour présenter leurs résultats et les méthodes développées. Dans le cadre de cette rencontre, un groupe de discussion a été réalisé pour explorer les expériences des co-évaluateurs et co-évaluatrices

du projet pilote ÉQUIPO, leur perception des forces et des faiblesses de cette méthode et les modifications à y apporter.

Encadré 3. Thèmes abordés en groupe de discussion lors du forum inter-organisationnel

1. Pouvez-vous décrire votre expérience de ce projet pilote ?
2. Quelles sont les forces et les faiblesses de la méthode ÉQUIPO ?
3. Quels défis avez-vous rencontrés ?
4. Comment cette méthode pourrait-elle être améliorée dans le futur ?
5. Qu'est-ce qui pourrait favoriser sa pérennité ?

Les prochaines sections présentent une analyse de ce groupe de discussion qui a également suivi un processus d'analyse thématique assistée par NVivo.

Répercussions du projet pilote ÉQUIPO

L'analyse thématique du groupe de discussion inter-organisationnelle a fait ressortir différents niveaux de répercussion de ce projet pilote (institutionnel, personnel, technique, professionnel, collaboratif).

Répercussions institutionnelles

Au niveau institutionnel, des participant-e-s ont noté que cette expérience a permis aux organisations d'apprendre de leurs forces et leurs faiblesses. Isis, psychologue au CJA, affirme que ce projet « a permis d'ajuster nos stratégies [...] de reconnaître ce que nous faisons bien et les choses que nous devons régler, penser et innover » (Isis, CJA). Cet apprentissage dans l'action s'est fait depuis les points de vue de bénéficiaires des ONG. Sonia, technicienne en ressources humaines au CPMGA, affirme avoir « pu approfondir et approcher la réalité vécue par les personnes pour découvrir nos forces et nos faiblesses » (Sonia, CPMGA). Toujours selon les participant-e-s, cet exercice réflexif aura permis d'aller au-delà de la contemplation des bons coups, pour analyser également les lacunes à combler. Christina, directrice du CPMGA, rappelle qu'il « est très facile de se regarder le nombril et nous dire que nous sommes bonnes (rire) [...], mais ce processus va au-delà [...] c'est aussi de voir nos faiblesses [...] pour moi la question

fondamentale de cet exercice est « Comment pouvons-nous nous améliorer ? » (Christina, CPMGA).

Ce projet pilote aurait offert aux parties prenantes une nouvelle façon d'évaluer les programmes. Selon Pilar, planificatrice de programmes au CPMGA, « cette expérience a permis d'explorer des espaces que nous n'explorions pas auparavant, [...] de connaître d'une autre manière comment nous touchons la population » (Pilar, CPMGA). L'évaluation qualitative, par l'approche inductive qu'elle préconise, aurait permis d'ajouter une perspective centrale sur les réalités évaluées. Calvin, psychologue au CJA, traite de l'importance « de connaître la réalité depuis le bas [population] vers le haut [institution] » (Calvin, CJA). Cette perspective inductive est un complément essentiel aux approches déductives généralement préconisées en planification et évaluation de programmes. L'évaluation qualitative amène une profondeur aux évaluations quantitatives généralement préconisées. Pilar affirme que l'une des erreurs des ONG est « de planifier les objectifs des programmes uniquement quantitativement [...]. Par exemple, avec l'indicateur « Nombre de femmes ayant participé à ce programme » [...], nous nous disons: Super ! Nous avons atteint notre objectif de couverture (rires) ! » (Pilar, CPMGA). Elle continue en affirmant que le projet a permis de compléter les chiffres en explorant :

> Comment les gens se sentent. Comment nous affectons leur vie. Lorsqu'on fait ce genre d'évaluation, nous pouvons nous approcher de leurs expériences [...]. Ceci nous amène plus loin. En les écoutant, vous vous rendez compte, au-delà des chiffres, de ce que vous faites bien ou mal (Pilar, CPMGA).

Les méthodes qualitatives ajoutent une dimension humaine aux indicateurs quantitatifs de suivi et d'évaluation préexistants. Selon la directrice du CPMGA, ce projet a enrichi « les indicateurs en terme qualitatif et ceci est très positif » (Christina, CPMGA). Jova tient des propos similaires : « Nous avions conçu de nombreux outils pour mesurer nos indicateurs quantitatifs, mais l'exercice que nous avons fait va au-delà de ce que nous avions jusqu'à présent avancé. Faire l'analyse du discours des gens, c'est une nouvelle expérience » (Jova, CJA).

L'expérience aurait également permis de clarifier la compréhension des méthodologies qualitatives. Jova affirme s'être rendue compte que l'organisation avait une fausse conception de ce qu'était le qualitatif :

> la formulation de nos indicateurs, que nous pensions être qualitatifs, soutenait plutôt une analyse quantitative [...]. Par exemple, le « pourcentage de femmes ayant expérimenté une hausse d'auto-estime ». Nous ne faisions aucune analyse du discours des femmes pour savoir comment elles avaient amélioré ou non leur auto-estime (Jova, CJA).

Répercussions personnelles

Au niveau personnel, ce projet aura sensibilisé les employé-e-s et les directions à l'importance de considérer les points de vue des bénéficiaires en interagissant directement avec elles et eux. Jova affirme que ce projet « a été très enrichissant pour les co-chercheuses et co-chercheurs, nous avons été grandement sensibilisés [...] il est beaucoup plus passionnant d'interagir avec les personnes de cette manière » (Jova, CJA). Toujours selon Jova, cette expérience aura été aussi motivante pour les bénéficiaires : « les jeunes filles et les femmes étaient très enjouées d'être partie prenante de ce projet ». Lorena du CJA spécifie qu'il « est important que les bénéficiaires soient des agentes dynamiques de l'évaluation [...] pas seulement par sondage. Elles ne sont pas que des bulletins de vote ! » (Lorena, CJA). L'aspect participatif d'ÉQUIPO concordait avec des valeurs partagées par la communauté et certains mouvements sociaux de la Bolivie contemporaine. Selon Rene, sociologue au CPMGA, « impliquer les gens dans le processus d'évaluation, concorde avec un point de vue postcolonial » (Rene, CPMGA).

Répercussions techniques

Au niveau technique, la mise en œuvre d'ÉQUIPO aura posé les bases d'un système d'organisation de l'information. Selon Lorena, « NVivo permet d'organiser l'information [...] nous devons continuer parce que nous avons d'autres idées qui ont émergé » (Lorena, CJA). Au-delà de l'analyse des données, NVivo peut être utilisé pour gérer, organiser et centraliser toutes les informations relatives à un projet ou à un programme offert par une organisation. Lorena affirme avoir hâte de « présenter à l'organisation cette nouvelle forme de voir le travail ». Ce type d'outil peut être utilisé pour faire le suivi à long terme des projets. Rene affirme qu'avec « NVivo, au-delà de

l'analyse, on peut documenter tout ce qui a été fait par l'organisation pour ainsi créer une mémoire plus riche » (Rene, CPMGA). Selon la directrice du CPMGA, « cela a été enrichissant de connaître un nouvel instrument de gestion de données » (Christina, CPMGA). À la fin du projet pilote, Christina a d'ailleurs initié un projet visant à centraliser et à analyser avec NVivo tous les rapports et la littérature produits par le CPMGA entre 2007 et 2012.

Répercussions professionnelles

Au niveau professionnel, malgré certains défis techniques, l'apprentissage des méthodes et d'un logiciel d'analyse qualitative a été un atout pour les co-évaluateurs et les co-évaluatrices, étant donné que peu de personnes maîtrisent ce genre d'outil dans le milieu du développement en Bolivie. Selon la directrice du CPMGA, « ce projet a permis aux techniciennes d'enrichir leur CV » (Christina, CPMGA). Lorena affirme à la blague : « J'ose dire qu'à Sucre nous sommes les quatre seules personnes qui savent comment utiliser NVivo (rires) » (Lorena, CJA). L'apprentissage de ce type de logiciel reste cependant un défi, spécialement pour ceux et celles n'ayant pas une grande dextérité informatique : « pour moi, la difficulté a été l'écart technologique, je me sentais complètement obsolète [...] je dois m'actualiser dans le maniement informatique » (Jova, CJA). Elle continue en soulignant l'importance du support d'un conseiller en analyse : « sans lui nous aurions été perdues » tout comme la rencontre de ses collègues du CPMGA [l'organisation partenaire], qui a permis d'échanger sur les défis rencontrés.

Répercussions collaboratives

Au niveau collaboratif, les échanges inter-programmatiques et inter-organisationnels auront été enrichissants pour les organisations en leur offrant une perspective globale sur les façons d'améliorer les programmes. Selon Christina, « la richesse fondamentale a été l'interaction entre les membres de l'équipe technique des différents programmes afin qu'elles puissent eux-mêmes constater comment nous pouvons nous améliorer » (Christina, CPMGA). Lorena continue en présentant l'importance de l'aspect inter-organisationnel d'ÉQUIPO qui aurait permis de mettre les bases d'une collaboration :

> l'une des choses les plus importantes a été de connaître les résultats de nos collègues du CPMGA [...] une institution que je connaissais de loin et je me suis rendue compte que nous avions beaucoup de choses en commun et je suis très enthousiaste à l'idée de continuer de partager nos méthodes.

Grâce à cet échange, les membres des organisations auront acquis une perspective complémentaire de la problématique de la violence envers les femmes dans une autre population [Quechua vs Aymara]. Cette expérience leur aura aussi permis de « mieux connaître les femmes d'El Alto [majorité Aymara] [...] pour comparer les choses que nous avons vues dans la population de Sucre [majorité Quechua], afin de pouvoir ainsi enrichir nos projets » (Lorena, CJA).

Forces, faiblesses et défis

L'importance de la mobilisation spontanée des employés

L'analyse thématique du groupe de discussion souligne que ce projet pilote a permis d'atteindre l'objectif principal de la méthode ÉQUIPO soit de « renforcer les capacités des partenaires dans l'utilisation des méthodes qualitatives informatisées et des approches participatives d'évaluation de programmes ». Plus spécifiquement, les sous-objectifs participatifs de cette méthode semblent avoir été atteints. En effet, des parties prenantes du CJA et du CPMGA ont été formées à la collecte et à l'analyse de données qualitatives assistée par ordinateur; des bases de données qualitatives pour explorer les expériences des bénéficiaires des ONG ont été générées; et un processus d'apprentissage dans l'action a été stimulé. Les sous-objectifs participatifs du projet ont permis de favoriser la participation active et démocratique des parties prenantes à titre de co-évaluateurs et co-évaluatrices de programmes des organisations partenaires. Suivant les sous-objectifs inter-organisationnels, la réalisation d'un forum a permis de créer un espace de partage pour promouvoir un échange entre le CJA et le CPMGA, deux ONG ayant des missions similaires. Cet événement a favorisé du même coup le partage d'expériences vécues lors de l'application de la méthode ÉQUIPO et renforcé la culture d'évaluation qualitative et participative des ONG partenaires.

L'une des forces de ce projet est le fait que celui-ci ait émergé d'une demande d'ONG locales. Des projets de collecte de données qualitatives étaient déjà en cours avant l'arrivée du conseiller. Par exemple, Isis avait réalisé une série d'entrevues pour évaluer l'impact d'un programme d'autonomisation des jeunes du CJA. Cependant, elle affirme qu'elle « ne savait pas quoi faire avec toute cette information, comment la transcrire et l'analyser [...] cela me frustrait beaucoup » (Isis, CJA). Elle continue en affirmant : « il était très important pour moi d'apprendre un outil de travail comme NVivo parce que cette recherche était en suspens ». L'accompagnement offert par le conseiller lui a permis d'analyser ces données et de publier un ouvrage (Barahona 2012).

La mobilisation spontanée des employé-e-s est un élément qui a favorisé la réalisation du projet au CJA. Une moins grande motivation a été remarquée chez les employé-e-s dont la participation a été « suggérée » par la direction du CPMGA. D'un autre côté, les employé-e-s de cette ONG ont bénéficié de plus de temps libre sur leur temps de travail, comparativement au CJA où les employé-e-s devaient souvent faire l'analyse dans leur temps libre. Un des défis partagés par les co-évaluateurs et co-évaluatrices du CJA a été de « mettre l'évaluation qualitative dans le plan opérationnel annuel [...] de prévoir du temps pour cela » (Isis, CJA).

Christina est convaincue de l'importance « de créer une culture institutionnelle d'évaluation qualitative et d'inclure cette méthode dans la planification annuelle, sinon elle sera reléguée au second plan [...]. Nous avons réalisé l'importance de ce type d'analyse, il faut donc l'inclure dans nos objectifs » (Christina, CPMGA). Sonia, technicienne en ressources humaines au CPMGA, a affirmé qu'elle allait « prévoir du temps dans les horaires des employé-e-s, des périodes prédéterminées où elles se verront, quelques fois par mois [...] je n'ai aucun doute qu'elles s'engageront, car c'est un outil très enrichissant pour leur travail, au quotidien » (Sonia, CPMGA). Jova rappelle cependant qu'un des enjeux est de « faire un suivi de l'application des recommandations qui sortent de cette évaluation. Comment les appliquer dans la mesure du possible ? [...] Comment les incorporer dans notre travail pour essayer d'améliorer nos faiblesses ? » (Jova, CJA).

Expériences et ressources des organisations

Le CJA et le CPMGA sont deux ONG bien établies employant plus de 40 personnes et offrant des services à des milliers de bénéficiaires (Tableau 1). Leurs programmes sont bien implantés et appuyés par différentes institutions de la coopération internationale. Si les ressources de ces organisations restent limitées, elles leur permettent néanmoins d'engager du personnel qualifié, d'accueillir des stagiaires universitaires et d'acheter du matériel nécessaire à l'évaluation qualitative (ordinateur et enregistreuse). Sans affirmer qu'ÉQUIPO ne peut être répliqué avec des ONG de moindre envergure, l'expérience et les ressources des organisations ont été deux facteurs ayant favorisé ce projet pilote.

Augmenter le niveau d'implication et d'engagement des bénéficiaires

Si les employé-e-s des ONG ont participé à l'ensemble du projet, les bénéficiaires « n'ont participé qu'à la réalisation et la transcription d'entrevues » (Jova, CJA). Selon Jova, l'un des défis sera de les faire participer à l'analyse de données. Pilar présente aussi l'enjeu de l'implication à long terme des bénéficiaires : « Comment faire que les gens participent depuis le début et restent ? [...] À la fin, on n'en avait plus beaucoup [...] Comment bien les engager dans le processus ? [...] Il faudra réfléchir à cela dans la planification » (Pilar, CPMGA). Pour Lorena, il est important de renforcer « la communication entre les co-évaluateurs et co-évaluatrices pour voir les autres perspectives, il faut ouvrir des espaces de partage et inclure des moments dans notre planification annuelle » (Lorena, CJA). Calvin ajoute qu'il « sera très stimulant d'être plus participatif, non seulement pour l'institution, mais aussi pour les bailleurs de fonds » (Calvin, CJA).

Être appuyé par une organisation qui applique elle-même la méthode ÉQUIPO

Si l'évaluation qualitative assistée par ordinateur était une nouvelle expérience pour le CJA et le CPMGA, elle l'était aussi pour Oxfam-Québec qui a montré l'exemple en appliquant ce processus à l'évaluation de ses propres programmes. En effet, la même série d'ateliers en méthodes qualitative et participative a été offerte à une douzaine d'agents et coopérants d'Oxfam-

Québec en 2012. Un vaste projet mobilisant la méthode ÉQUIPO a été réalisé au Pérou et en Bolivie. 17 groupes de discussions, rejoignant 144 bénéficiaires et employé-e-s de 12 ONG locales [dont CJA et CPMGA], ont été réalisés et analysés en équipe pour explorer les forces et les faiblesses des programmes d'OXFAM-Québec auprès des partenaires. Les co-évaluateurs et co-évaluatrices de CJA et CPMGA ont noté avoir apprécié cet effort d'Oxfam-Québec qui montre l'exemple en appliquant lui-même ce qu'il suggère aux partenaires. Selon Jova, ceci est un aspect qui démarquait Oxfam-Québec : « Au niveau de la coopération internationale, vous [Oxfam] avez incorporé ce type d'analyse qualitative, et ceci m'apparait excellent, mais les autres institutions de la coopération ne sont pas rendues là » (Jova, CJA).

Convaincre les bailleurs de fonds de la plus-value des méthodes qualitatives

Un défi majeur de l'application de l'évaluation qualitative en contexte de développement international est de convaincre les bailleurs de fonds qui privilégient la mesure d'indicateurs quantitatifs. Selon Jova, la plupart d'entre eux :

> demandent un suivi d'indicateurs quantitatifs uniquement [...] Quand on présente du qualitatif, ils nous disent « faite un effort pour quantifier, pour voir la couverture ». Nous avançons dans une direction, mais la coopération est dans une autre direction, on doit s'y plier, suivre le modèle logique, les indicateurs quantitatifs [...] Réellement, obtenir une concession pour le qualitatif s'est difficile (Jova, CJA).

Pilar tient des propos similaires : « Ils veulent voir les effets sur leurs indicateurs quantitatifs » (Pilar, CPMGA).

Au-delà de l'évaluation, il semblerait que les méthodes qualitatives ne soient pas privilégiées par les bailleurs de fonds lors de la planification de programmes. Selon Calvin, ceci expliquerait l'inefficacité des plusieurs projets qui ne sont pas adaptés aux réalités des terrains et des besoins des populations cibles :

> plusieurs projets ne fonctionnent pas parce qu'on a mal évalué le contexte [...] la coopération internationale nous arrive avec des projets déjà élaborés qui ne sont pas adaptés à la réalité. Je crois

> qu'avec un outil comme NVivo nous pouvons analyser ce que veut la population cible, c'est une bonne façon de contextualiser. (Calvin, CJA)

Après avoir constaté dans l'action la valeur ajoutée du qualitatif, les membres des ONG locales doivent s'attarder à convaincre à l'externe. Selon Rene, « nous savons que le qualitatif permet de voir des choses qui ne se voient pas avec le quantitatif, mais le problème est à l'externe où le qualitatif n'est pas reconnu [...] il est important de convaincre sinon nos efforts seront vains » (Rene, CPMGA). Selon Jova, l'une des stratégies pour promouvoir les méthodes qualitatives dans un contexte dominé par les chiffres serait de les présenter en complément au quantitatif : « les aspects qualitatifs vont être d'abord présentés en complément au quantitatif » (Jova, CJA).

Analyse réflexive et conclusion : vers les méthodes mixtes

Ce projet pilote d'ÉQUIPO a permis de renforcer les capacités et l'intérêt des partenaires envers l'utilisation des méthodes qualitatives et participatives d'évaluation de programmes. L'apprentissage d'un logiciel d'analyse qualitative a permis de construire une base de données qualitative qui pourra être alimentée pour suivre l'application des leçons apprises. Cependant, comme l'exprime Jova à la fin de la section précédente, il faut considérer que l'auditoire auquel s'adresse les évaluateurs et les évaluatrices est davantage sensibilisé aux approches quantitatives que qualitatives. Présenter d'abord le qualitatif en complémentarité du quantitatif, comme le propose Jova, est une stratégie d'introduction qui a été utilisée dans d'autres domaines, notamment en sciences de la santé qui ont été historiquement fortement dominées par les approches quantitatives (Creswell et Plano Clark 2011). Pour continuer le travail sur l'aspect inter-organisationnel, il serait intéressant que les deux ONG puissent développer une base de données complémentaire intégrant leurs données quantitatives et qualitatives.

Pilar souligne la complémentarité entre l'évaluation « d'indicateurs quantitatifs qui nous montrent, par exemple, combien de personnes ont complété le programme, et les indicateurs qualitatifs qui nous permettent d'explorer leur réalité, par exemple, si elles sont retombées en situation de violence et pourquoi. Puis, comment pourrions-nous continuer à les aider ? » (Pilar, CPMGA). Rene met l'accent sur l'apport potentiel des méthodes mixtes

en évaluation : « Un regard quantitatif nous permet d'explorer les impacts, alors que les processus peuvent se comprendre davantage d'un point de vue qualitatif. [...] Le meilleur serait d'avoir une perspective mixte » (Rene, CPMGA).

Les méthodes mixtes sont de plus en plus valorisées en évaluation de programmes (Bamberger 2012; Patton 2015; Bujold, El Sherif *et al.* 2018). L'analyse thématique assistée est également utilisée pour faire la méta-analyse de rapports d'évaluations de programme de développement (Bujold 2017). L'une des lacunes de ce projet a été son manque d'intégration de l'analyse qualitative de données avec les évaluations quantitatives des programmes. Une prochaine étape d'ÉQUIPO pourrait être l'Évaluation Qualitative-Quantitative Intégrative et Participative Organisationnelle (EQQUIPO) intégrant l'analyse de données quantitatives et qualitatives dans un même projet d'évaluation. L'enseignement et l'apprentissage des méthodes mixtes comportent cependant d'autres défis qui devront être pris en compte (Bujold, Hong *et al.* 2018).

L'établissement d'une communauté de pratique interdisciplinaire et interculturelle est une stratégie pour surpasser les défis des méthodes mixtes en évaluation. C'est pour cette raison que nous proposons à tous les lecteurs et lectrices intéressées par les méthodes mixtes de devenir membre de l'association Méthodes mixtes Francophonie qui vise la collaboration entre des chercheuses, chercheurs et praticien-ne-s francophones de différentes aires géographiques et culturelles.

Remerciements

Les auteurs remercient Oxfam-Québec et ses représentants, le CPMGA et le CJA ainsi que les co-évaluateurs et co-évaluatrices de ce projet pilote, les coopérant-e-s volontaires et tou-te-s les bénéficiaires qui ont participé. Ce projet a été rendu possible grâce à l'appui du Gouvernement du Canada (Affaires Mondiales Canada) qui finançait le programme « Mieux agir et mieux influencer » exécuté par Oxfam-Québec.

Références clés

Patton, M. Q. (2015). *Qualitative research et evaluation methods: integrating theory and practice.*Thousand Oaks, California : SAGE Publications.

Une ressource précieuse pour explorer différents aspects et applications des méthodes de recherche et d'évaluations qualitatives.

Miles, M. B., Huberman A. M. et Saldaña J. (2014). *Qualitative data analysis : a methods sourcebook. Lieu d'édition : Maison d'édition.*

Un incontournable pour approfondir les méthodes d'analyse de données qualitatives.

Bujold, M., Hong, Q. N., Ridde, V., Bourque, C. J., Dogba, M. J., Vedel, I. et Pluye, P. (2018). *Oser les défis des méthodes mixtes en sciences sociales et sciences de la santé.* Montréal : 117e Cahiers scientifiques de l'ACFAS.

Premier livre en français et en accès libre sur les méthodes mixtes. Disponible gratuitement à l'adresse suivante : http://methodesmixtesfrancophonie.pbworks.com.

Pour en savoir plus sur les ONG et les associations citées dans ce chapitre et leurs programmes de développement, les lecteurs et les lectrices peuvent consulter les liens suivants :

- http://centrojuanaazurduy.org
- http://gregorias.org.bo/
- http://methodesmixtesfrancophonie.pbworks.com
- http://oxfam.qc.ca/

Références

Bamberger, M. (2012). *Introduction aux méthodes mixtes dans l'évaluation d'impact.* Repéré sur le site d'InterAction : https://www.interaction.org/document/introduction-aux-méthodes-mixtes-dans-l'évaluation-d'impact

Barahona, I. (2012). *Young labor right: A study of young migrants from rural 1st and 2nd generation*. Edited by Juanas Azurduy Center. Sucre, Bolivia.

Bazeley, P. et Jackson, K. (2013). *Qualitative Data Analysis with NVivo* (éd.). Colorado: Sage Publications.

Bujold, M. (2014). *Méthodologies de recherche et d'évaluation qualitatives : de la théorie à la pratique, guide de formation*. Montréal, Canada : CAQI.

Bujold, M. (2016). *NVivo: un support à l'analyse qualitative, guide de formation*. Montréal, Canada : CAQI.

Bujold, M. (2017). *Analyse thématique et critique de rapports initiaux d'évaluation des portefeuilles pays d'ONU Femmes : réalisations, défis et recommandations à considérer*. New York : ONU Femmes.

Bujold, M., El Sherif, R., Bush, P. L., Johnson-Lafleur, J., Doray, G. et Pluye, P. (2018). Ecological content validation of the information assessment method for parents (IAM-parent): A mixed methods study. *Evaluation and Program Planning*, 66, 79-88.
https://doi.org/10.1016/j.evalprogplan.2017.09.011

Bujold, M., Hong, Q. N., Ridde, V., Bourque, C. J., Dogba, M. J., Vedel, I. et Pluye, P. (2018). *Oser les défis des méthodes mixtes en sciences sociales et sciences de la santé*. Montréal : 117e Cahiers scientifiques de l'ACFAS.
http://methodesmixtesfrancophonie.pbworks.com

Cargo, M. et Mercer, S. L. (2008). The value and challenges of participatory research: strengthening its practice. *Annual Review of Public Health*, 29, 325-50.
doi: 10.1146/annurev.publhealth.29.091307.083824

Creswell, J. W. et Clark, V. L. P. (2011). *Designing and conducting mixed methods research* (éd.). Los Angeles : Sage Publications.

Daigneault, P.-M. et Jacob, S. (2009). Toward accurate measurement of participation: Rethinking the conceptualization and operationalization of participatory evaluation. *American Journal of Evaluation*, 30(3), 330-348.

Fereday, J. et Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods*, 5(1), 80-92.

Jagosh, J., Macaulay, A. C., Pluye, P., Salsberg, J., Bush, P. L., Henderson, ... Greenhalgh, T. (2012). Uncovering the benefits of participatory research: implications of a realist review for health research and practice. *The*

Milbank Quarterly, 90(2), 311-46.
doi: 10.1111/j.1468-0009.2012.00665.x

Laperrière, A. (1997). Les critères de scientificité des méthodes qualitatives. Dans J.-P. Deslauriers, J. Poupart, L.-H. Groulx, A. Laperrière, R. Mayer et A.P. Pires (dir.), *La recherche qualitative : enjeux épistémologiques et méthodologiques* (p. 365-390). Boucherville, QC : Gaëtan Morin.

Miles, M. B., Huberman, A. M. et Saldaña, J. (2014). *Qualitative data analysis : a methods sourcebook*. Lieu d'édition : Maison d'édition

Ministerio de justicia. (2015). *Vidas dignas, mujeres libres de violencia*. Bolivie : Éditeur.

Olivier de Sardan, J.-P. (2008). *La rigueur du qualitatif : les contraintes empiriques de l'interprétation socio-anthropologique*. Louvain-la-Neuve, BE: Academia-Bruylant.

Paillé, P et Mucchielli, A. (2016). *L'analyse qualitative en sciences humaines et sociales* (4e éd.). Paris : Armand Colin.

Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. *Health services research*, 34(5 Pt 2), 1189-1208.

Patton, M. Q. (2015). *Qualitative research et evaluation methods : integrating theory and practice*. Thousand Oaks, CA : Sage Publications.

Plante, J. (1994). *Évaluation de programmes*. Sainte-Foy, QC : Presses de l'Université Laval.

Porteous, N. L. (2012). La construction du modèle logique d'un programme. Dans V. Ridde et C. Dagenais (dir.), *Approches et pratiques en évaluation de programmes* (p. 89-107). Montréal : Les Presses de l'Université de Montréal.

Ridde, V. et Dagenais, C. (2012a). Introduction générale à l'évaluation de programmes. Dans V. Ridde et C. Dagenais (dir.), *Approches et pratiques en évaluation de programmes* (p. 13-32). Montréal : Les Presses de l'Université de Montréal.

Ridde, V. et Dagenais, C. (2012b). *Approches et pratiques en évaluation de programmes : nouvelle édition revue et augmentée*. Montréal : Les Presses de l'Université de Montréal.

Résumé / Resumen / Abstract

Les programmes de développement visent à produire des changements dans des communautés touchées par des problématiques socioculturelles complexes, par exemple l'égalité entre les femmes et les hommes, où s'entrecroise une pluralité de significations. L'évaluation de ces programmes nécessite une approche systématique mobilisant la collaboration entre les parties prenantes de différentes organisations. Les trois volets d'Évaluation QUalitative Informatisée et Participative inter-organisationnelle (ÉQUIPO) visent à répondre à ces réalités complexes. Le volet qualitatif vise à comprendre les processus de changement depuis les points de vue de ceux ou celles qui expérimentent une situation complexe. Le volet participatif en favorisant la participation démocratique des parties prenantes rappelle que l'évaluation de programmes n'est pas une fin en soi, mais un moyen d'améliorer les conditions sociales et collectives. Le volet inter-organisationnel stimule les échanges entre des ONG partageant des missions similaires. Cette méthode d'apprentissage dans l'action a été développée en 2012, lors d'un projet pilote d'évaluation de programmes d'ONG boliviennes promouvant l'égalité entre les femmes et les hommes. L'objectif de ce chapitre n'est pas de présenter les résultats de l'évaluation spécifique des programmes, mais plutôt de décrire la méthode ÉQUIPO, le contexte dans laquelle elle a émergé, et de présenter ses répercussions et les défis de son application depuis les perspectives des co-évaluateurs et évaluatrices qui ont participé à ce projet pilote.

***

Los programas de desarrollo tienen como objetivo, producir cambios en comunidades con problemas socioculturales complejos, por ejemplo la desigualdad de las mujeres, donde se entrecruzan una pluralidad de significaciones. La evaluación de estos programas, necesita de un enfoque sistemático que movilice la colaboración entre las partes interesadas de diferentes organizaciones. Los tres componentes de la Evaluación Cualitativa Informatizada y Participativa inter-Organizacional (ÉQUIPO), apuntan a responder a estas realidades complejas. El objetivo del componente cualitativo es comprender los procesos de cambio desde puntos de vista diversos de aquellos o aquellas que experimentan una situación compleja. El componente participativo está orientado a promover la participación democrática de las partes interesadas y recuerda que la evaluación de

programas no es un fin en sí, sino un medio para mejorar las condiciones sociales y colectivas. El componente inter-organizacional estimula los intercambios entre las ONG compartiendo misiones similares. Este método de aprendizaje en la acción, ha sido desarrollado en el 2012 durante un proyecto piloto de evaluación de programas de ONG bolivianas, promoviendo la igualdad entre hombres y mujeres. El objetivo de este capítulo no es presentar los resultados de la evaluación específica de programas, sino más bien describir el método ÉQUIPO, el contexto en el cual ha emergido, y la presentación de repercusiones y desafíos de su aplicación desde las perspectivas de los co-evaluadores que participaron en este proyecto piloto.

***

Development programmes aim to produce change in communities affected by complex socio-cultural issues, such as gender equality, where a plurality of meanings intersect. Evaluating these programmes requires a systematic approach that involves collaboration between stakeholders from different organizations. The three components of the QUalitative Informative and Inter-Organisational Participatory Assessment (EQUIPO) aim to respond to these complex realities. The qualitative component aims to understand change processes from the perspectives of those experiencing a complex situation. The participatory component, by promoting the democratic participation of stakeholders, reminds us that programme evaluation is not an end in itself, but a means of improving social and collective conditions. The inter-organizational component stimulates exchanges between NGOs sharing similar missions. This method of learning in action was developed in 2012, during a pilot project to evaluate programmes of Bolivian NGOs promoting gender equality. The purpose of this chapter is not to present the results of the specific program evaluation, but rather to describe the TEAM methodology, the context in which it emerged, and to present its implications and the challenges of its application from the perspectives of the co-evaluators who participated in this pilot project.

***

Mathieu Bujold, détenteur d'un doctorat en anthropologie, est chercheur postdoctoral au Département de médecine de famille de l'Université McGill. Ses expériences de recherche, son cursus universitaire

et son parcours professionnel sont à l'intersection des sciences sociales, des sciences de la santé, des sciences de l'éducation et de l'évaluation de programme. Depuis 2007, il offre ses services de formation en analyse qualitative informatisée et méthodes mixtes à plusieurs étudiant-e-s, chercheuses et chercheurs et professionnels et a été expert invité dans différents départements universitaires, associations de recherche et organisations internationales œuvrant en développement (par exemple la Banque Mondiale). Outre le développement méthodologique, ses intérêts de recherche portent sur le pluralisme médical, les soins de santé intégrés, la collaboration interprofessionnelle en santé et les soins centrés sur la personne.

Jean-Alexandre Fortin, détenteur d'une maîtrise en Aide humanitaire internationale, est Coordonnateur du Programme Accès et Innovation d'Oxfam Québec en Bolivie. Ses expériences de plus de dix ans en Bolivie avec Oxfam-Québec lui ont permis de connaître en profondeur le contexte sociopolitique de pays. Son cursus universitaire et son parcours professionnel sont spécialisés en gestion de programmes de développement auprès des femmes et des jeunes,particulièrement en ce qui a trait aux aspects de planification, suivi et évaluation de programmes. Depuis 2003, il a travaillé avec différentes organisations non gouvernementales œuvrant en développement international tel que le Centre de Coopération Internationale en Santé et Développement, Jeunesse Canada Monde et Oxfam-Québec.

Citation

Mathieu Bujold et Jean-Alexandre Fortin. (2019). L'évaluation qualitative, informatisée, participative et inter-organisationnelle (EQUIPO). Un exemple d'un programme en faveur des femmes victimes de violences en Bolivie. In *Évaluation des interventions de santé mondiale. Méthodes avancées*. Sous la direction de Valéry Ridde et Christian Dagenais, pp. 57-82. Québec : Éditions science et bien commun et Marseille : IRD Éditions.

4. La méthode photovoix

Une intervention auprès de populations marginalisées sur l'accès à l'eau potable, l'hygiène et l'assainissement au Mexique

LYNDA REY, WILFRIED AFFODÉGON, ISABELLE VIENS, HIND FATHALLAH ET MARIA JOSÉ ARAUZ

La méthode photovoix

Définition de la méthode

« La photovoix est un processus par lequel les individus peuvent identifier, représenter et améliorer leur communauté à travers l'utilisation d'une technique photographique spécifique, en confiant des caméras à des personnes pour leur permettre d'agir comme enregistreurs et catalyseurs potentiels du changement dans leurs propres communautés » (Wang et Burris, 1997, p. 369, traduction libre)

Forces de la méthode

- Méthodologie flexible et facilement adaptable à divers contextes, objectifs et audiences
- Matériel visuel (photos) comme tremplin pour la réflexion personnelle et comme canal pour s'exprimer facilement
- Permet une atmosphère d'échange et de confiance
- Fait appel à la responsabilité individuelle et collective des participant-e-s face à des enjeux identifiés dans leur communauté
- Stimule la créativité des participant-e-s
- Favorise la liberté d'expression : des commentaires positifs et moins positifs sont librement exprimés sur le projet

Défis de la méthode

- Nécessite une importante préparation : disponibilité des ressources techniques (impression de grandes photographies de qualité et de format adéquats), espace physique adéquat, équipement.
- Disponibilité de ressources humaines formées et ayant de l'expérience avec des jeunes et des populations vulnérables
- Compétences des facilitateurs et des facilitatrices de l'activité : compréhension de la méthode, ingéniosité, flexibilité, grande réceptivité et écoute
- Risque d'influence parfois non intentionnelle de certains participant-e-s leaders durant l'activité, au détriment des autres participant-e-s
- Présence d'un traducteur ou d'une traductrice capable de ne pas rompre la fluidité de la communication et la confiance de tous les participant-e-s dans un contexte où la langue maternelle des participant-e-s est autre que celle de l'animateur ou de l'animatrice de l'exercice

La photovoix (*photovoice*) est une méthode qualitative de collecte de données souvent utilisée en recherche-action, à travers laquelle les participant-e-s prennent des photos pour explorer des sujets complexes ou sensibles[1]. Créée au début des années 1990, elle est basée sur l'idée que les photos peuvent aider les participant-e-s, souvent socialement en marge, à raconter une histoire, identifier les problèmes auxquels ils et elles font face et réfléchir à la formulation de leurs propres solutions. Ainsi, la photographie ne sert pas uniquement à répondre à des questions de recherche descriptives : elle est également un moyen d'amorcer un véritable changement de paradigme (Fradet 2012).

> *Photovoice is a process by which people can identify, represent, and enhance their community through a specific photographic technique. It entrusts cameras to the hands of people to enable them to act as recorders, and potential catalysts for change, in their own communities* (Wang et Burris 1997, p. 369).
>
> La photovoix est un processus par lequel les individus peuvent identifier, représenter et améliorer leur communauté à travers une technique photographique spécifique, en confiant des caméras à des personnes pour leur permettre d'agir comme enregistreurs et catalyseurs potentiels du changement dans leurs propres communautés (Traduction libre).

À quoi sert la photovoix ?

La photovoix, méthode visuelle et participative, vise en général trois objectifs principaux : 1) permettre aux individus de mettre en évidence les forces et les préoccupations de leur communauté; 2) promouvoir la connaissance et le dialogue critique en discutant en groupe de questions importantes; 3) influencer les décisions politiques en mettant en évidence des problématiques sociales (Wang et Burris 1997; Wang 1999; Wang et Redwood-Jones 2001; Fradet 2012).

En effet, à travers la photo se révèle *le pouvoir de l'image* dans la définition du réel et du normal. L'image permet d'interpréter la réalité qui

1. Le projet dont l'exemple nourrit ce texte a été financé par la Fondation One Drop que nous remercions.

nous entoure. Elle agit également comme un véhicule des émotions suscitées par des conditions économiques et socioculturelles qui affectent les individus et leurs communautés. Le processus de la photovoix est très mobilisateur : c'est un espace favorisant la prise de conscience des réalités sociales et où a lieu la création de représentations symboliques et de sens. Les autres peuvent ainsi voir le monde à travers les yeux du ou de la photographe si bien que des perspectives et interprétations multiples cohabitent. Grâce au processus et aux sujets sociaux qui y participent activement, la photovoix a le potentiel d'agir comme catalyseur de changement social.

De la photovoix au photolangage : quels lieux communs ?

Il existe plusieurs variantes de la photovoix dont une appelée « photolangage ». La photovoix se différencie du photolangage au niveau de leurs fondements. Alors que la photovoix s'inscrit dans une visée conscientisante, émancipatrice et de changement social, le photolangage repose à l'origine sur une approche plus thérapeutique. Créé en 1965 par un groupe de psychologues français qui travaillaient avec des adolescent-e-s, le photolangage a été conçu de manière intuitive. L'objectif était d'utiliser des photos pour servir de soutien à la parole de jeunes qui rencontraient des difficultés à s'exprimer et à parler en groupe de leurs expériences diverses et parfois douloureuses sur le plan personnel. Il s'agit également d'une approche spécifique de travail de groupe à partir d'une collection de photos choisies pour illustrer un thème particulier. Le photolangage se concentre davantage sur la prise de conscience individuelle, centrée sur des problèmes personnels, voire psychologiques, et sur une prise de parole devant l'ensemble des participant-e-s (Vacheret, 2011).

Une autre différence importante existe entre les deux techniques au niveau de leurs processus respectifs. En effet, dans le cas du photolangage, les photos sont prises par un ou une photographe et la sélection préalable de photos est faite par le facilitateur ou la facilitatrice de l'activité, contrairement à la démarche photovoix où les participant-e-s prennent eux-mêmes les photos (White, Sasser, Bogren et Morgan 2009). Toutefois, les différences s'estompent lorsque la démarche du photolangage s'aligne sur les fondements théoriques et les concepts de la conscientisation comme instrument libérateur dans un processus d'éducation et de transformation

sociale (Freire, 1970). La seule différence qui subsiste est alors au niveau de l'auteur ou l'autrice des photos (le ou la photographe vs les participant-e-s).

Le potentiel d'influence de l'image/photo utilisée comme « porte-voix familier, pouvoir, véhicule d'émotions, indicateur et révélateur symbolique des réalités sociales, catalyseur de changement et outil de plaidoyer et d'action sociale » (Fradet, 2012) est reconnu dans les deux approches.

Comme processus, la photovoix et le photolangage jouent un rôle dans l'éveil de la « conscience critique », c'est-à-dire une compréhension profonde de la façon dont le monde fonctionne et comment la société, la politique et les relations de pouvoir affectent la réalité d'un individu ou d'un groupe. La technique de la photographie, en « dévoilant » les réalités sociales et en les exposant, peut influencer l'émergence des changements sociaux (Danchin et Sax 2009; White *et al.* 2009).

La photovoix, une méthode de recherche et d'évaluative participative

La photovoix est généralement présentée comme une méthode de la recherche-action participative, mais elle peut également être utilisée en tant que méthode de recherche qualitative ou comme un outil de collecte de données pour la planification et l'évaluation (White *et al.*, 2009). Axée sur la communauté, cette méthode est flexible et peut servir différents objectifs (mobilisation communautaire, évaluation des besoins, évaluation de processus et des résultats dans une perspective participative) (Catalani et Minkler 2010; Wang et Burris 1997). Mise en œuvre dans le cadre de la planification d'une intervention et de l'évaluation des besoins, le processus de photovoix implique la communauté afin de mieux comprendre leurs réalités, notamment, à travers le point de vue de personnes qui ne sont pas en position de pouvoir ou d'autorité. Il s'agit d'une forme de légitimation des connaissances communautaires (Mayfield-Johnson, Rachal et Butler 2014; Strack, Lovelace, Jordan et Holmes 2010).

Utilisée comme méthode d'évaluation participative, la photovoix permet d'avoir accès à des informations concrètes et utiles sur les communautés ciblées (Goo-Kuratani et Lai 2011). Elle peut être complétée par plusieurs méthodes telles que les enquêtes sur les ménages, la technique du groupe nominal et les groupes de discussion tout en équilibrant les objectifs de recherche, d'action et d'évaluation. Généralement, à partir de discussions

entre les participant-e-s et les membres de l'équipe de recherche, une question est identifiée. Par la suite, les participant-e-s munis de leurs caméras partent en quête de photographies qui permettront d'y répondre (Wang et Burris 1997). Dans une étude de Wang *et al.* (1996), la photovoix a été utilisée avec des femmes chinoises en zone rurale impliquées dans un projet de santé communautaire pour évaluer de façon participative l'efficacité, les forces et les faiblesses du projet. La façon dont la photovoix peut contribuer à la capacité d'une communauté à réfléchir sur sa propre culture est discutée. La photovoix a également été utilisée par Duffy (2010) dans le cadre d'une évaluation formative où le chercheur a questionné des femmes seules sur les bénéfices de participer à un projet de recherche. La photovoix est d'autant plus porteuse lorsqu'elle est mise en œuvre en collaboration avec des populations vulnérables ou marginalisées telles que les femmes, les jeunes, les personnes souffrant d'une maladie particulière, les sans-abris (Valiquette-Tessier *et al.* 2015) ou encore les travailleuses du sexe (Desyllas, 2014).

Le processus de photovoix

En règle générale, la mise en œuvre de la photovoix se déroule en trois phases. Elle nécessite la présence d'un facilitateur ou d'une facilitatrice qui connait bien les réalités historiques, politiques et socioculturelles du milieu d'intervention. *La première phase* consiste en la formation et le renforcement des capacités des participant-e-s qui se familiarisent avec les outils (caméras, batteries, pellicules, projecteurs, albums, etc.). La *deuxième phase* concerne la collecte de données, ce qui correspond aux discussions à l'intérieur de groupes de personnes (préalablement identifiées) sur les enjeux de développement communautaire. Enfin, la *dernière phase* est l'analyse qualitative de propos/discours des participant-e-s à travers lesquels elles et ils mettent en relation les photographies avec leurs réalités (Catalani et Minkler 2010; Wang et Burris 1997).

Avantages et défis liés à la photovoix

Palibroda *et al.* (2009) ont identifié plusieurs avantages et défis relatifs à l'utilisation de la photovoix, notamment au niveau des participant-e-s,

du processus de recherche et de facilitation ainsi qu'au niveau de la communauté. Rappelons que la mise en œuvre de la photovoix favorise notamment (i) le renforcement de l'engagement de la communauté dans l'action et le plaidoyer, (ii) une meilleure compréhension des besoins de la communauté, et (iii) l'augmentation de l'autonomisation (*empowerment*) individuelle et collective (Catalani et Minkler 2010).

L'approche comporte également plusieurs défis. D'abord, les leaders d'opinion influents dans les communautés défavorisées peuvent tirer profit de leur statut pour imposer leur point de vue et maintenir un *statu quo* face à des enjeux de développement ou sociaux, au détriment de la majorité marginalisée. Les participant-e-s pourraient aussi avoir du mal à présenter une réalité complexe ou abstraite à travers leurs photographies.

Ensuite, la photovoix étant centrée sur l'individu, elle peut aussi mettre davantage en valeur les points de vue individuels plutôt que le consensus communautaire. Lorsque les rapports de pouvoir sont en faveur de certains individus au détriment d'autres, la photovoix pourrait perpétuer les inégalités sociales. La discussion autour de questions parfois sensibles peut induire des sentiments négatifs.

Enfin, vivant souvent en situation de vulnérabilité, les communautés/individus peuvent avoir d'autres priorités que celles d'entrer dans une telle démarche, outre le manque de temps. Les résultats de la photovoix pourraient aussi ne pas être à la hauteur des attentes des membres de la communauté, surtout en ce qui a trait au changement social.

Le tableau suivant présente les avantages et les limites de la démarche photovoix.

Tableau 1. Résumé des avantages et défis de la méthode Photovoix selon Wang et Burris (1997)

Avantages	**Défis**
• Approche descriptive à travers une forme de communication par l'image • Favorise l'interaction entre les participant-e-s à travers les idées, les histoires, etc. • Valorise les connaissances des communautés • Encourage la participation communautaire • Méthode adaptée aux communautés vulnérables et marginalisées • Peut être utilisée comme outil d'identification des besoins et méthode d'évaluation • Approche de renforcement de capacités, d'apprentissage et d'autonomisation permettant de prioriser les besoins et de trouver des solutions.	• Apprentissage de la prise de photos peut être difficile • Complexité de l'analyse des données recueillies durant les discussions • Ne permet de couvrir qu'un nombre limité d'enjeux sociaux • Décalage entre la méthode et la pratique : nécessité de s'adapter au contexte d'application et à l'audience en temps réel • Risque de maintien du *statu quo* par les leaders d'opinion • Met davantage l'accent sur le jugement personnel au détriment du consensus communautaire. • Peut perpétuer les inégalités sociales en accordant plus de voix aux lobbies

Dans la seconde partie de ce chapitre, nous présentons une expérience de photovoix utilisée comme méthode d'évaluation formative qui se rapproche davantage du photolangage décrit plus haut. En effet, dans le cas présenté ici, les photos ont été prises par les membres de l'équipe de projet et non par les participant-e-s en raison certains paramètres (méthode non planifiée en début de projet, mais ajustement agile en cours de mise en œuvre permettant une meilleure adaptation au contexte et aux participant-e-s, la contrainte de temps et au besoin d'obtenir des données quasi en temps réel pour prendre des mesures correctives avant le déploiement des activités en phase 2 du projet). Cette expérience est d'autant plus intéressante que la photovoix y est utilisée comme méthode d'évaluation basée sur les arts (*Art-Based Evaluation*) (Desyllas 2014). Elle apporte une « autre » façon de porter un jugement de valeur sur une intervention en utilisant des outils et approches créatives[2]. L'objectif de la démarche était d'évaluer le volet artistique d'un projet d'accès à l'eau, l'hygiène et

2. Art-based evaluation 101 prepared for ArtReach Toronto by Margo Charlton of Resonance Creative Consulting Partners. G.O.A.L. YOUTH Workshop Series. http://www.artreachtoronto.ca/toolkits

l'assainissement mis en œuvre dans une communauté autochtone du Mexique.

La photovoix pour évaluer l'art social : entre complexité et créativité

Contexte

Dans un contexte où 700 millions de personnes n'ont toujours pas accès à l'eau potable et 80 % d'entre elles vivent en milieu rural (WHO/UNICEF 2015), la Fondation One Drop basée à Montréal, en partenariat avec une fondation privée et une organisation non gouvernementale mexicaines, a lancé en 2015 un projet pilote visant un accès durable à l'eau potable, l'hygiène et l'assainissement dans deux municipalités (Mixtla de Altamirano et Tehuipango) de l'État de Veracruz au Mexique.

Les deux municipalités sont peuplées par des communautés indigènes partageant la même langue (98 % des habitants parlent le nahuatl). Elles sont caractérisées par un fort taux d'analphabétisme et un manque d'infrastructures en eau potable et assainissement. Pendant une longue période de l'année, les infrastructures en eau sont non fonctionnelles à cause notamment d'un coût d'entretien très élevé. Pour satisfaire leurs besoins en eau, les populations s'approvisionnent dans les eaux de rivière et autres sources non potables. Certaines familles possèdent des réservoirs dans lesquels elles peuvent stocker de l'eau de pluie, mais la quantité collectée demeure insuffisante.

Par ailleurs, bien que la majorité des ménages dispose de latrines (environ 96 % en Mixtla et 86 % en Tehuipango selon les données du projet), ces infrastructures sont inappropriées en raison du manque d'hygiène et de sécurité.

En matière d'hygiène personnelle, l'absence, la faible fréquence, ainsi qu'une pratique inappropriée du lavage des mains et du corps avec du savon ainsi que du brossage des dents ont été observées.

L'ensemble de ces pratiques, combinées au fait qu'une majorité de ménages ne traitent pas l'eau avant de la boire, favorisent la prolifération de germes pathogènes et les maladies hydriques au sein des communautés et des familles.

Tableau 2. Contexte des communautés ciblées (INEGI 2010; *One Drop* 2015)

Municipalités ciblées (Communautés)	**Tehuipango** (Apitpitzactitla)	**Mixtlade Altamirano** (Mixtlantlapak)
Taux d'analphabétisme	Supérieur à 50 %	Supérieur à 50 %
Langue	Nahuatl (à 98 %)	Nahuatl (à 98 %)
Population cible	23 746	10 387
Taux de couverture en assainissement	86 %	96 %
Taux de couverture en eau	41 %	55 %
Hygiène	Comportements d'hygiène inadéquats ayant pour conséquence une haute incidence des maladies liées à l'eau	

La logique de l'intervention

Face à ces problématiques structurelles et comportementales, un projet mis en œuvre par l'ONG mexicaine a financé la construction d'infrastructures hydrauliques et d'assainissement. Toutefois, les infrastructures seules ne suffisent pas pour encourager l'adoption de bonnes pratiques et d'habitudes au niveau individuel et communautaire. La Fondation One Drop et ses partenaires ont donc décidé d'intégrer dans le projet d'infrastructure un volet centré sur l'art social : une approche systématique utilisant toutes formes d'art pour aborder des enjeux sociaux à travers la sensibilisation et la mobilisation sociale. One Drop mise sur le potentiel de l'art social comme moteur de transformation sociale et de changement de comportements dans les projets en eau, hygiène et assainissement (EHA) qu'elle soutient (www.onedrop.org).

Coin de l'hygiène au sein de la classe. Crédit photo : One Drop, 2016

Opérationnalisé en collaboration avec les artistes locaux, l'art social permet aux individus de participer activement et de manière créative à la résolution des problèmes vécus au sein de leur communauté. L'intervention par les arts conduit au développement et à la concrétisation d'actions individuelles et/ou collectives répondant aux besoins des personnes et des communautés. Les participant-e-s qui s'impliquent dans ces projets artistiques se responsabilisent, acquièrent de nouvelles habiletés en prenant part à des expériences enrichissantes tout en développant leur estime d'eux-mêmes et d'elles-mêmes. Plusieurs acteurs et actrices de développement y ont eu recours depuis des décennies, notamment dans le domaine de la santé et tout particulièrement du VIH/Sida (Onuekwe 2015; Moyer-Gusé 2008; Singhal et Rogers 1999). L'art social se situe au croisement de plusieurs stratégies de communication pour le développement, d'éducation populaire et tout particulièrement de l'approche d'éducation-divertissement (*edu-entertainment*).

Au Mexique, les objectifs visés par la composante Art Social pour le Changement de Comportement du projet étaient de 1) intégrer l'art social comme outil de sensibilisation, mobilisation et de changement des comportements individuels et collectifs dans un projet EHA existant et 2) évaluer les effets de l'art social sur les connaissances, attitudes et pratiques autour de deux comportements cibles : le lavage des mains et le traitement de l'eau. De façon spécifique, une telle démarche visait, d'une part, à concevoir et à intégrer les activités d'art social dans deux communautés cibles et, d'autre part, à développer un cadre intégré de suivi, d'évaluation et d'apprentissage avec des méthodes innovantes permettant d'évaluer les effets de l'art social sur les variables d'intérêt mentionnées précédemment.

Ci-dessous se trouve le modèle logique de l'intervention en art social, c'est-à-dire les activités prévues et les résultats attendus à court, moyen et long terme, en tenant compte du contexte du projet et de l'environnement des communautés cibles.

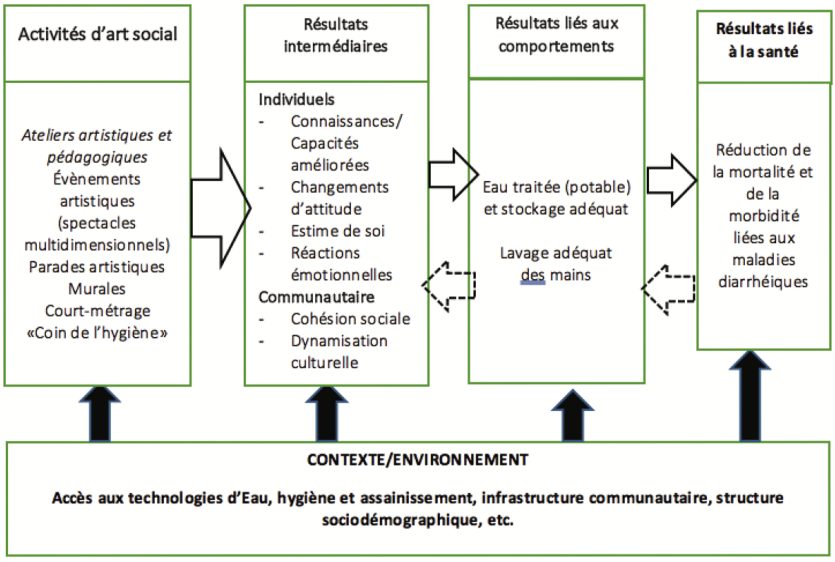


Figure 1. Modèle logique du projet (inspiré de Figueroa; Kincaid 2010)

Peintures murales sur les réservoirs d'eau. Crédit photo : One Drop, 2016

Dans le cadre de cette initiative novatrice, des ateliers d'art social incluant le théâtre, la danse (contemporaine et traditionnelle), la musique, le cirque, les arts visuels et la réalisation de courts-métrages ont été organisés auprès des enfants et des adolescent-e-s dans les deux communautés du projet. Les thèmes abordés par les artistes dans ces ateliers ainsi que les messages véhiculés étaient centrés sur le lavage des mains et le traitement de l'eau. Le processus de création et les formes artistiques étaient largement inspirés des récits, contes et autres atouts socioculturels locaux. Les ateliers reposaient sur des processus de co-création collective permettant d'améliorer les attitudes, les connaissances et les pratiques des participant-

e-s autour du lavage des mains et du traitement de l'eau tout en leur apprenant des techniques artistiques.

Selon ce processus, il était attendu qu'à l'issue de leur expérience d'éveil artistique, d'apprentissage cognitif et de réflexion critique, les capacités/connaissances des jeunes en tant qu'agents et agentes de changement au sein de leurs communautés seraient renforcées, qu'ils et elles gagneraient en estime de soi et que les attitudes par rapport aux problématiques visées seraient positives et proactives en vue de favoriser un changement de comportement durable au niveau des individus et de la collectivité.

Création du coin de l'hygiène avec les instituteurs/trices. Crédit photo : One Drop, 2016

En complément des ateliers artistiques susmentionnés, l'activité « coin de l'hygiène » a été mise en œuvre. Elle a consisté en la création d'un espace dédié à l'hygiène corporelle dans les salles de classe. L'aménagement de ce coin a été réalisé dans le cadre d'ateliers ludiques et éducatifs de bricolage, peinture et collage auxquels ont participé élèves et enseignant-e-s. Cette activité été mise en œuvre afin d'encourager la créativité des enseignant-e-s dans le développement de matériels éducatifs permettant de sensibiliser les enfants des écoles primaires aux enjeux liés à l'hygiène et d'adopter des pratiques adéquates dans un environnement ludique et attrayant.

D'autres événements ciblant les communautés de manière plus large ont été organisés dans le cadre du projet : des parades de danse, des spectacles multidisciplinaires, la diffusion des courts-métrages produits dans le cadre des ateliers et des peintures murales.

Le tableau suivant présente des activités mises en œuvre, les groupes cibles, les participant-e-s, et les thématiques abordées.

Tableau 3. Activités d'art social implantées dans les zones ciblées

Activités implantées	**Description et participant-e-s**	**Groupes cibles/ audiences**	**Thématiques ciblées**
Ateliers artistiques et pédagogiques	Théâtre, danse contemporaine et traditionnelle, musique, cirque, arts visuels, réalisation de courts-métrages	Jeunes de 12 à 18 ans	Apprentissages techniques et cognitifs dans plusieurs domaines artistiques intégrant un contenu en lien avec les pratiques de lavage des mains et de traitement de l'eau au niveau du ménage
Événements artistiques (spectacles)	Création, production et présentation de spectacles multidisciplinaires par les jeunes accompagnés par les artistes	Communauté entière	Pratiques de lavage des mains et de traitement de l'eau au niveau du ménage
Parades artistiques	Création et organisation par les jeunes accompagnés par les artistes/Participation active des jeunes, des artistes et de membres de la communauté au défilé final : une parade culturelle basée sur des expériences de caravane traditionnelle	Communauté entière	
Murales	Peintures réalisées par les jeunes soutenus par les artistes/Inauguration par les jeunes pendant le défilé final	Communauté entière	
Courts-métrages	Création par les jeunes des ateliers formés par les artistes/Présentation par les jeunes pendant le défilé final	Communauté entière	

« Coin de l'hygiène »	Participation de tous les élèves des classes concernées	Élèves du primaire (6-12 ans)	Pratiques de lavage des mains et d'hygiène personnelle des enfants

Le plan d'évaluation des activités d'art social

Diffusion des courts-métrages. Crédit photo : One Drop, 2016

Dans le cadre de ce projet, plusieurs outils de suivi et d'évaluation quantitatifs et qualitatifs ont été développés et utilisés, particulièrement pour mesurer les effets des interventions artistiques. La méthode de la photovoix a été adaptée[3] et appliquée en tant qu'outil d'évaluation auprès des jeunes participant aux ateliers d'art social afin d'obtenir leurs opinions, perceptions et leur satisfaction relativement à ces ateliers.

Un plan d'évaluation a été développé en collaboration avec l'équipe de projet des zones ciblées. Il s'agit d'un plan qui aborde plusieurs questions, mais seules celles qui étaient visées par la photovoix sont exposées ici.

3. Comme dans le cas du photolangage, les photos ont été prises par l'équipe de projet, mais les objectifs visaient la conscientisation des jeunes et ont encouragé leur analyse critique quant à leur participation aux différents ateliers, comme ce qui est préconisé dans le cas d'une démarche de photovoix.

Tableau 4. Plan d'évaluation de la composante art social du projet

Objectif de l'évaluation	**Questions d'évaluation (élaborées en collaboration avec l'équipe de projet et les artistes)**	**Processus et outils de collecte de données**
Évaluer les perceptions des jeunes relativement aux activités artistiques auxquelles ils et elles avaient participé dans le cadre de la composante Art Social pour le Changement de Comportement en lien avec l'eau potable et l'hygiène (traitement de l'eau et lavage des mains)	RÉACTIONS Quelles ont été les réactions des participant-e-s aux activités d'art social?APPRENTISSAGES Quelles ont été les leçons apprises par les jeunes à travers leur participation aux activités d'art social?EFFETS SUR LES PRATIQUES ET COMPORTEMENTS Quelle a été l'influence de l'art social sur les pratiques liées à l'eau et à l'hygiène telles que le lavage des mains et le traitement adéquat de l'eau afin de réduire le risque de maladies?RÉSULTATS Quels sont les effets proximaux de l'art social sur la communauté?	Photos prises par l'équipe de projet et les artistes suivant des critères bien établis et identifiés par les jeunes Discussions de groupes Dessins réalisés par les jeunes pour exprimer leurs réactions et perceptions Entrevues individuelles avec les jeunes leaders (non considérées dans ce texte) Enquêtes des ménages sur les connaissances, attitudes et pratiques (CAP) relatifs à l'eau, l'hygiène et l'assainissement (non considérées dans ce texte)

Application de la photovoix comme démarche évaluative en contexte artistique et social

Objectifs

La photovoix a été choisie comme approche d'évaluation formative. La démarche avait l'avantage d'être adaptée au contexte indigène dans lequel le projet a été mis en œuvre, au caractère artistique des activités et aux jeunes participant-e-s. L'activité de photovoix visait deux objectifs :

- Partager les représentations des participant-e-s sur les questions liées à l'eau et à l'hygiène, exposer les idées reçues, identifier les stéréotypes et aller au-delà en donnant la parole à l'expérience personnelle de chaque individu.
- Permettre aux participant-e-s de s'exprimer librement et de libérer leurs émotions, individuellement et en groupe, sur le sujet d'intérêt tout en ayant une réflexion critique sur les solutions à envisager.

Processus

La phase de planification

Le processus de photovoix nécessite des préparatifs qui impliquent du temps et des ressources humaines. Pour assurer la mise en œuvre efficace du processus d'évaluation sur le terrain, une coordinatrice de projet spécialisée en art social et en évaluation, soutenue par l'équipe de projet (une personne chargée de la traduction, une personne chargée de la prise de notes durant l'exercice et un photographe/caméraman), a été placée au sein de la communauté. Ces personnes ont été formées et orientées par l'évaluatrice interne de One Drop basée à Montréal qui a réalisé plusieurs déplacements sur le terrain pour soutenir la mise en œuvre de l'évaluation.

Les principales étapes de planification nécessaires à la réalisation de l'évaluation sont les suivantes.

Sélection des participant-e-s

Le choix des participant-e-s à la démarche évaluative par la photovoix s'est réalisé selon les critères suivants :

- Participation active des jeunes aux ateliers artistiques décrites plus haut : ceux et celles qui ont été assidus ont été priorisés afin d'obtenir une vision complète des effets de l'art social
- Choix paritaire de filles et de garçons
- Volonté à s'exprimer devant un groupe et consentement éclairé à participer à la photovoix
- Capacité à comprendre le nahuatl et/ou l'espagnol

Les groupes étaient composés de huit participant-e-s au minimum et de 15 au maximum, ce qui constitue un nombre relativement adéquat pour ce type d'activité de groupe. Une fois la sélection terminée, la coordonnatrice du projet a communiqué la liste des participant-e-s aux leaders de chaque communauté ainsi qu'aux parents afin de pouvoir obtenir leur consentement à réunir les participant-e-s identifiés par la sélection. Le consentement libre

des parents et des enfants a été obtenu et les informations clés liées à l'activité artistique (objectif, date, heure, lieu) ont ensuite été communiquées aux participant-e-s. Une salle de classe a été mise à disposition par la municipalité pour l'activité.

Sélection des photos

Avant l'activité de photovoix, les artistes et l'équipe de projet ont sélectionné les photos qui ont servi de support pour l'évaluation selon les critères suivants :

Parade au sein de la communauté. Crédit photo : One Drop, 2016

- Photos représentant des scènes et des participant-e-s en action
- Photos représentant chacun des ateliers artistiques facilités par les artistes dans le cadre du projet art social
- Photos représentant les thèmes du projet : lavage des mains et traitement de l'eau
- Photos reflétant les émotions des jeunes participant au projet durant les ateliers.

Une vingtaine de photos ont ainsi été sélectionnées en tenant compte du nombre de participant-e-s (environ 10). Pour chacune des photos choisies, deux copies ont été réalisées, notamment dans le cas où deux participant-e-s souhaiteraient sélectionner la même photo. Chacune des photos a été identifiée par un numéro au verso pour faciliter l'analyse. Enfin, les photos ont été imprimées sur un grand format (8 ½ x 11) afin de permettre aux participant-e-s de mieux visualiser les détails de chacune des photos.

Préparation des ressources humaines

La question des ressources humaines est essentielle dans cette activité. Il est utile de disposer de plusieurs personnes-ressources pouvant appuyer

l'activité (traduction, prise de notes, etc.). Toutefois, elles doivent être orientées sur l'attitude à adopter pendant la démarche afin que leur présence n'entraîne pas de gêne de la part des participant-e-s.

La photovoix a été menée dans deux communautés ciblées par le projet et quatre personnes-ressources ont à chaque fois assumé les rôles suivants :

- Animation
- Observation/Prise de notes durant l'exercice
- Traduction
- Photographie/tournage

Ces personnes avaient le mandat d'appuyer le bon déroulement de l'exercice. Elles ont donc reçu préalablement une courte formation couvrant les points suivants : 1) explication du processus de la photovoix; 2) présentation du processus et des étapes de réalisation; 3) explication des enjeux de l'exercice; 4) identification et clarification des rôles de chacun-e durant la photovoix

.

Spectacle multidisciplinaires avec les jeunes. Crédit photo : One Drop, 2016

Dans le cadre de ce projet, la coordonnatrice des activités artistiques sur le terrain a été choisie pour être l'animatrice de la photovoix. L'objectif était de capitaliser sur la confiance et la complicité existant entre elle et les jeunes, et d'encourager leur participation active. En effet, la proximité de la coordonnatrice avec les jeunes a permis de créer un climat de confiance si bien que les jeunes se sont sentis à l'aise pour s'exprimer librement. L'évaluatrice interne de One Drop était présente lors de l'exercice en tant qu'observatrice. Elle a supervisé l'activité et observé le déroulement ainsi que l'attitude non-verbale des jeunes. Elle a aussi orienté l'animatrice sur certains aspects, notamment s'assurer que tous les participant-e-s prennent la parole; préciser certaines questions; donner le temps aux participant-e-s plus timides de s'exprimer. Les artistes ayant facilité les ateliers artistiques n'étaient pas présents afin de

permettre aux jeunes de s'exprimer sur leur expérience lors de ces ateliers de manière ouverte et en toute confiance.

La mise en œuvre

L'activité de photovoix menée dans chacune des deux communautés s'est déroulée comme suit :

Organisation de l'espace du local (30 minutes)

L'organisation du local compte beaucoup dans la création d'une atmosphère de partage et d'écoute. Des chaises ont été disposées dans la salle de classe en cercle pour encourager l'échange, l'interaction et l'écoute active. Outre l'espace de discussion circulaire, une grande table a été installée au fond de la salle et les photos sélectionnées y ont été exposées. La salle de classe a été un choix judicieux, car les jeunes étaient dans un espace familier et avaient à disposition du matériel scolaire.

Disposition des jeunes en cercle favorisant le dialogue, crédit photo : One Drop 2016

Présentation du travail au groupe (15 minutes)

Avec l'appui constant du traducteur, l'animatrice s'est chargée d'expliquer et de détailler de façon claire et précise les étapes du processus à suivre, les méthodes utilisées (notamment l'importance de l'écoute et la possibilité de s'exprimer librement) et les contraintes horaires. L'objectif était de s'assurer que chaque participant-e ait compris la démarche et la méthodologie, et se sente en confiance pour participer activement à l'activité. Cette introduction a également intégré la présentation des différentes personnes présentes dans la salle et de leur rôle, afin de ne pas créer de malaise chez les jeunes.

Choix individuel des photographies (10 minutes)

Le choix individuel des photographies devait permettre aux participant-e-s de s'exprimer autour des questions de perceptions suivantes :

- *Question 1 : Choisissez une photo qui décrit ce que vous avez le plus aimé dans les ateliers*
- *Question 2 : Choisissez une photo qui décrit ce que vous avez le moins aimé dans les ateliers*

Pour cette étape, l'animatrice a invité les jeunes à se déplacer vers la table où étaient exposées les photographies pour procéder de manière individuelle au choix de la photo qui décrivait le mieux ce qu'ils avaient apprécié dans les ateliers (Question 1). Tout en se déplaçant autour de la table de manière silencieuse, ils et elles ont identifié une photo en relation avec la première question. Puis, l'animatrice a invité les participant-e-s à prendre la photo choisie et à revenir dans l'espace d'échange.

Pour celles et ceux qui avaient choisi la même photo, des copies étaient disponibles afin que toutes et tous aient en main leur photo pour la discussion.

Après avoir complété l'échange (décrit à l'étape suivante) autour de la première question, le même procédé a été utilisé pour la question numéro 2 : quelle est la photo qui représente ce que vous avez le moins aimé dans les ateliers?

Jeunes choisissant les photos représentant la réponse à la question posée, crédit photo : One Drop 2016

Travail de groupe sur les questions 1 et 2 (durée : 1 heure)

Après le choix de la photo en lien avec la première question, le cercle de discussion s'est formé afin de pouvoir échanger sur les réactions et réflexions de chacun des participant-e-s.

L'animatrice a répété la question et a invité chaque jeune à partager sa réflexion en lien avec la question en présentant au groupe la photo choisie, les raisons de ce choix et la réponse à la question. Les autres participant-e-s étaient encouragés à adopter une attitude d'écoute active et à attendre leur tour de parole afin de ne pas influencer les réflexions de chaque jeunes présent à l'exercice.

De plus, l'animatrice a rappelé les objectifs de l'exercice afin que chacun-e des participant-e-s se sente à l'aise de s'exprimer.

Chaque participant-e était libre de s'exprimer en espagnol ou en nahuatl. Selon la langue choisie, les propos ont été traduits en temps réel. Chacun des participant-e-s a pu bénéficier de son temps de parole et tous ont manifesté le désir de communiquer leurs émotions quant au choix de leur photo. Il n'y a pas eu de réticences ni de jeunes qui ne souhaitaient pas s'exprimer.

Lorsqu'approprié, l'animatrice a posé des questions supplémentaires durant les échanges afin de pousser plus loin la réflexion de chacun des jeunes.

Ce même travail de réflexion a été réalisé pour la deuxième question.

Jeunes s'exprimant sur les photos choisies, crédit photo : One Drop 2016

Travail individuel sur la question 3 (durée : 30 minutes)

Cette phase a fait appel à la créativité des jeunes et à leur capacité à « dessiner leurs émotions ». Dans un processus traditionnel de photovoix ou de photolangage, le dessin n'est pas utilisé. L'équipe de projet a décidé d'inclure la composante dessin dans l'évaluation par la photovoix pour deux raisons : 1) il était intéressant d'intégrer une autre méthode de collecte de données basée sur l'art et 2) les arts visuels ayant une place de choix dans la culture indigène, le dessin favoriserait l'expression libre et différente (non verbale) des jeunes pour exprimer leurs émotions.

Durant cette phase, les participant-e-s devaient répondre à la question suivante :

Question 3 : Pouvez-vous faire un dessin qui décrit comment vous vous sentez depuis cette expérience ?

Cette question avait pour objectif d'amener les jeunes à exprimer leurs émotions après la fin des ateliers et des autres activités d'art social qui ont eu lieu dans leur communauté. Au lieu d'un échange en groupe comme pour les questions précédentes, les jeunes se sont isolés dans le lieu de leur choix, cour d'école, autre salle, autour d'une table, pour réaliser leur dessin.

Du matériel de dessin a donc été fourni et un temps d'environ 30 minutes accordé à chaque participant-e pour la réalisation des dessins. Ces derniers pouvaient être accompagnés par des mots ou des phrases. Les jeunes ont exploré la question de manière individuelle en respectant leurs espaces de travail et leur individualité respective.

Jeunes réalisant leurs dessins, crédit photo : One Drop 2016

Certains ont repris des idées qu'ils avaient exprimées durant les questions 1 et 2 tandis que d'autres ont exprimé de nouvelles idées et produit des images fortes et symboliques.

Conclusion de l'activité (10 minutes)

Jeune revenant sur son expérience, crédit photo : One Drop 2016

L'exercice d'évaluation par la photovoix s'est conclu par la présentation de manière volontaire des dessins réalisés durant l'exercice précédent. Les participant-e-s étaient invités à partager leurs réflexions à partir de leur création.

Ensuite, l'animatrice a remercié et félicité les jeunes pour leur participation active et leur respect des consignes. Les jeunes ont également eu la possibilité de partager leurs impressions sur

l'activité et ont beaucoup apprécié revenir sur leur expérience dans les ateliers artistiques de manière ludique et artistique.

Au total, le processus de photovoix aura duré environ deux heures. Les participant-e-s ont pu retourner à la maison et les membres de l'équipe ont procédé à un bilan de l'activité et partagé leurs impressions sur son déroulement. Il a également été décidé de produire des copies des photographies choisies par les participant-e-s afin de les leur offrir. Pour renforcer le sentiment de fierté dans leurs communautés respectives, les photos des jeunes participant-e-s ont été exposées et rendues visibles par tous.

L'analyse des données

Les données collectées lors d'une démarche de photovoix sont principalement qualitatives et le processus d'analyse peut présenter des défis. Dans le cadre de ce projet, il s'agissait de collecter les perceptions et les réactions des jeunes sur leur expérience durant les ateliers artistiques et pédagogiques. Les questions posées aux jeunes étaient les suivantes :

Question 1 : Choisissez une photo qui décrit ce que vous avez le plus aimé dans les ateliers

Question 2 : Choisissez une photo qui décrit ce que vous avez le moins aimé dans les ateliers

Question 3 : Pouvez-vous faire un dessin qui décrit comment vous vous sentez depuis cette expérience ?

Les stratégies d'analyse utilisées sont présentées ci-dessous :

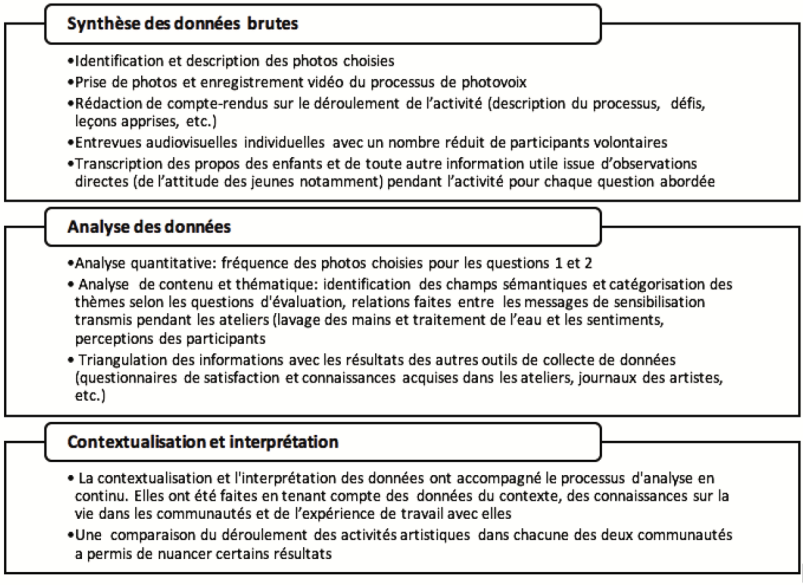


Figure 2. Processus d'analyse des données collectées

Résultats de l'évaluation

Réactions des participant-e-s aux activités d'art social du projet

Jeunes en fin d'exercice de photovoice, crédit photo : One Drop 2016

La majorité des jeunes ont déclaré avoir apprécié **l'atmosphère des ateliers** (le divertissement, le travail de groupe et la complicité avec leurs camarades). Toutefois, certains ont aussi soulevé des situations conflictuelles vécues dans certains ateliers et d'autres ont rapporté avoir beaucoup moins apprécié l'atelier musique, moins divertissant selon eux. Ceci s'expliquerait par le manque d'expérience des animateurs et animatrices avec les jeunes autochtones et qui n'ont pas réussi à motiver les enfants.

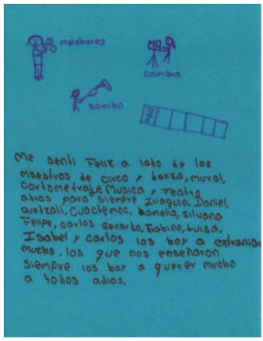
La relation avec les artistes (attachement, complicité, confiance) a été particulièrement appréciée, notamment en raison de leur accessibilité, de leur adaptabilité à l'audience et de leur habileté à motiver, divertir, mobiliser et sensibiliser les jeunes. Selon Mackinnon et Ryan (2015), les activités artistiques comme la photovoix sont aussi une opportunité pour les animateurs et animatrices de démontrer leurs habiletés de leadership au-delà de l'objectif initial.

Tout en faisant partie d'un processus créatif collectif, le potentiel d'épanouissement individuel des jeunes est accru à travers la **reconnaissance du succès collectif** : « Ce qui m'a le plus touché, c'est de voir mes amis réussir des acrobaties qu'ils n'arrivaient pas à faire au début de l'activité » (jeune participant).

Les photos des jeunes participant-e-s aux ateliers, crédit photo : One Drop 2016

Les jeunes ont ensuite exprimé un **sentiment d'abandon et de tristesse** à la fin des activités (départ des artistes notamment, perte d'accès au matériel artistique, rupture du groupe créé pendant les ateliers). Somme toute, la photo, le dessin et la parole ont permis aux jeunes d'exprimer leurs émotions les plus profondes et parfois mitigées.

Apprentissages des participant-e-s aux ateliers d'art social



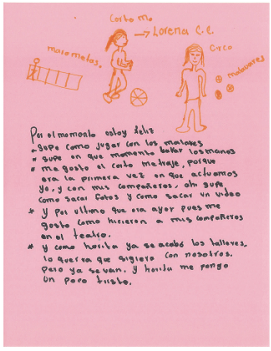
« Je me sentais heureuse aux côtés des formateurs de cirque et danse, peinture murale, court-métrage, musique et théâtre. Au revoir à jamais Joaquin, Daniel, Quetzali, Cuactemoc, Pamelia, Silvana, Felipe, Carlos Gerardo, Gabino, Luisa, Isabel et Carlos [prénoms des artistes]. Vous allez beaucoup me manquer. Vous qui nous avez appris. Je vais toujours vous aimer beaucoup. Au revoir à tous » (traduction libre). Crédit photo : One Drop 2016

Cette activité a confirmé que l'un des effets des ateliers artistiques et pédagogiques est le **renforcement des capacités personnelles des jeunes et la libération de leur plein potentiel.** Les résultats de plusieurs projets financés par One Drop en Afrique (Burkina Faso), en Inde (Odisha) et en Amérique centrale (El Salvador et Honduras) et impliquant les jeunes dans des activités d'art social décrivent des processus visant à faire des jeunes des agents de changement au sein de leurs communautés. Lors de ces ateliers, les jeunes, et notamment les filles, avaient démontré un leadership individuel et ont progressivement gagné en confiance entre le début et la fin des activités, ce qui a été confirmé par l'expérience mexicaine.

Le sentiment de fierté et d'efficacité personnelle (*self-efficacy*) s'accroit notamment à la suite de l'apprentissage de nouvelles techniques artistiques. Cette attitude positive favorise l'assimilation des apprentissages de pratiques saines telles que le lavage des mains et le traitement de l'eau pour la rendre potable. Lors des activités, les jeunes ont pu pratiquer à plusieurs reprises un lavage des mains adéquat. Toutefois, la durabilité de ces apprentissages peut être remise en cause s'ils ne sont pas répétés dans leur quotidienneté. Pour atténuer ces risques, certaines activités ont été développées en étroite collaboration avec les instituteurs et institutrices afin de les intégrer au curriculum scolaire et de continuer à les implanter au sein de leur classe. Dans la plupart des cas, des parents étaient présents lors des activités.

Perceptions des participant-e-s à la photovoix à propos de la démarche évaluative

La photovoix a été un succès dans les deux communautés. Les participant-e-s ont fait preuve d'une grande confiance et d'une ouverture face à l'exercice malgré la présence d'intervenants externes (caméraman, traducteur). Ils et elles ont également été très respectueux des consignes (silence pendant le choix, écoute, etc.), ce qui a permis de créer une atmosphère d'échange et de dialogue. Ce constat est accord avec les résultats de plusieurs études (Delgado 2015; Rania *et al.* 2014).



« En ce moment je suis content : je savais comment jouer avec les balles de jonglage. Je savais à quels moments me laver les mains. J'ai aimé le court-métrage, parce que ce fut la première fois que nous y participions moi et mes amis ; maintenant je sais comment prendre des photos et la façon de faire une vidéo. Et enfin, hier comment mes amis et moi avons fait du théâtre. Et aujourd'hui comme les ateliers sont terminés, j'aurais voulu continuer, et maintenant je me sens un peu triste. » (traduction libre). Crédit photo : One Drop 2016

> Comme la photovoix est orientée vers les participant-e-s, les questions préliminaires qui guident une étude sont malléables, ce qui permet une plus grande réactivité aux besoins identifiés des participant-e-s, ainsi que l'exploration de nouveaux domaines de découverte au fur et à mesure qu'ils émergent (traduction libre; Patton 2002, cité par Sutton-Brown 2014, p. 171[4]).

Considérant la grande timidité de certains participant-e-s et tenant compte du défi de la langue maternelle (nahuatl) de ces communautés, la photovoix a facilité la communication des idées à travers le médium visuel comme support à la parole. Les jeunes ont été particulièrement réceptifs aux photographies. Ces dernières étaient un centre de curiosité et d'interaction, ainsi que le point de départ d'une réflexion cognitive : les photos ont permis aux jeunes de se replonger dans le souvenir des moments passés pendant les ateliers et

4. As photovoice is participant-directed, the preliminary questions that guide a study are malleable. This openness allows for increased responsiveness to the participants' identified needs, as well as for the exploration of new areas of discovery as they emerge.

d'amorcer une analyse critique de leur expérience (Strack *et al.* 2004). La photo sélectionnée joue le rôle du « masque » comme au théâtre, permet d'être soi derrière une image qui nous représente et à travers laquelle les sentiments sont exprimés. La photo est aussi « un miroir de l'âme » offrant aux jeunes de se percevoir en action pour le changement. Le choix de la photovoix pour porter un regard rétrospectif sur les activités d'art social a été jugé pertinent, car elle a permis d'encourager la réflexion personnelle du participant ou de la participante sur les effets que ces activités ont eus sur son individualité et sa place au sein du groupe. La prise de conscience et la responsabilité personnelle sur des enjeux sociaux (dans ce cas-ci des enjeux de santé), dans un contexte où la notion d'appartenance communautaire prime sur l'individualité, ont été favorisées. La photovoix a révélé l'autonomisation des participant-e-s en leur offrant l'espace nécessaire pour être eux-mêmes et elles-mêmes, de développer et exprimer des opinions personnelles dans leur interaction avec le groupe dans une atmosphère respectueuse de partage et d'écoute.

La photovoix s'est avérée ludique et originale. Les jeunes ont tous été très enthousiastes à y participer (Madrigal *et al.* 2014; Strack *et al.* 2004). Contrairement à des formes d'évaluation plus traditionnelles (enquêtes, focus groupes) auxquelles elles et ils ont été confrontés tout au long de la mise en œuvre du projet, les participant-e-s ont déclaré s'être amusés, divertis et avoir été touchés émotionnellement. Ils ont pu s'exprimer de façon créative sans avoir l'impression d'*évaluer* les activités auxquelles ils et elles avaient participé ou *d'être évalués*. Dans leur étude, Madrigal *et al.* constatent que l'intégration d'activités artistiques dans un processus de photovoix aide à créer une expérience intéressante et divertissante pour les jeunes tout en cultivant leur esprit d'équipe et leur sens de la communauté.

Les participant-e-s ont de la confiance et de l'assurance pour exprimer leurs idées, que ce soit concernant les aspects positifs comme ceux à améliorer dans les ateliers. Les jeunes ont également affirmé leur individualité et démontré un leadership collectif durant l'exercice tel qu'identifié par Fradet (2012) à travers l'acronyme VOICE : **V***oicing* **O***ur* **I***ndividual and* **C***ollective* **E***xperience*. Ils ont en effet été en mesure de choisir personnellement, sans l'influence de leurs compagnons participant-e-s à l'activité, leurs photos, et d'émettre des opinions consensuelles et parfois différentes, y compris sur une photo identique. Malgré les barrières de la

langue parfois, les jeunes se sont exprimés librement, sans hésitation et sans inhibition.

Analyse réflexive sur la photovoix comme outil d'évaluation

La photovoix a été utilisée pour la première fois par One Drop au Mexique. Le projet d'art social centré sur le changement de comportements a constitué une opportunité de mettre en œuvre des approches d'évaluation innovantes et qualitatives permettant de recueillir les perceptions de participant-e-s issus de communautés indigènes et de mieux comprendre les effets de l'art social, difficilement quantifiables et mesurables à l'aide d'outils traditionnels.

L'art social est une intervention complexe, fortement dépendante de son contexte d'implantation. En ce sens, l'utilisation de la photovoix est appropriée, car elle s'adapte à la culture et au contexte dans lequel vivent les participant-e-s. Par exemple, elle ne nécessite pas de traduction de questionnaire en langue locale et fait appel à la créativité des jeunes tout en favorisant leur libre expression. Cette première expérience a permis d'identifier les forces et les aspects à améliorer dans la conduite de la démarche; nous avons pu capitaliser sur ces apprentissages.

Il faut souligner aussi que la décision d'appliquer la photovoix de façon agile et adaptée a permis d'obtenir des informations utiles et utilisables qui ont contribué à mobiliser et sensibiliser la communauté (qui a rarement l'occasion d'accueillir des activités ludiques pour les jeunes), à améliorer les activités subséquentes dans d'autres projets, à partager les pistes d'amélioration avec les partenaires notamment les artistes et à intégrer la photovoix comme activité à part entière dans le modèle d'intervention en Art Social pour le Changement de Comportement de l'organisation.

Leçons apprises

La sélection et l'impression des photos : une étape importante à planifier

Lors du premier exercice de photovoix dans la première communauté,

les dimensions requises (photos suffisamment larges pour faciliter la sélection par les participant-e-s) n'avaient pas été respectées. Il a donc fallu faire preuve d'ingéniosité et réimprimer des photos prises tout au long du projet avec une taille suffisamment grande pour plus de visibilité. Ce contretemps a permis de révéler l'importance de bien planifier une étape aussi anodine que l'impression des photos, notamment en double et bien à l'avance. Les dimensions des photos sont essentielles puisque la photo constitue l'élément principal de l'activité. Plus les photos sont grandes, plus elles permettent aux participant-e-s de comprendre la scène et de se remémorer les moments décrits ou de s'identifier aux photos. Par ailleurs, la diversité des photos et leur nombre (au minimum 6 photos, maximum 10) sont également importants, afin d'offrir un large choix aux participant-e-s et de refléter le plus d'opinions possible.

La langue : un facteur essentiel de participation et d'implication

Le projet a été implanté au sein de communautés parlant le nahuatl, une langue autochtone du Mexique. Dans ce contexte, les interventions de terrain, incluant les collectes de données, devaient prendre en compte la complexité de la communication. Toutes les activités nécessitant de recueillir les perceptions, les commentaires ou les suggestions des populations se réalisaient par conséquent en nahuatl. Cela permettait une plus grande fluidité et mettait les participant-e-s en confiance. Il a été observé à plusieurs reprises que les personnes participant au projet avaient beaucoup de plus facilité à s'exprimer dans leur langue maternelle qu'en espagnol.

Ainsi, durant l'exercice de la photovoix, le choix a donc été laissé aux participant-e-s de s'exprimer en nahuatl ou en espagnol. Les participant-e-s ont décidé en grande majorité de s'exprimer en nahuatl. Étant donné qu'une partie des facilitateurs et facilitatrices de l'exercice ne maîtrisaient pas le nahuatl, une participante a volontairement procédé à la traduction durant l'intervention. De ce fait, cette participante n'a pas été en mesure de participer pleinement à l'activité. Pour éviter cette situation, un traducteur indépendant a été engagé pour procéder à la traduction simultanée durant l'exercice de photovoix dans la deuxième communauté. Wang et Burris (1997) suggèrent que la photovoix soit adaptée aux besoins des participant-e-s de façon flexible. La question de la langue de communication choisie pour

mener l'activité a donc été essentielle dans le cadre de la photovoix et des autres activités menées dans cette région du Mexique. Afin que les participant-e-s comprennent l'exercice et participent pleinement, en étant en mesure de s'exprimer facilement, il est indispensable de créer un environnement favorable à leur implication. Ceci suppose de prendre en compte la langue dans laquelle les participant-e-s sont le plus à l'aise pour s'exprimer et de planifier un système de traduction simultanée si nécessaire pour favoriser les interactions. Le fait de pouvoir parler sa propre langue augmente la qualité de données collectées. Les émotions s'expriment plus facilement en langue maternelle, et donc l'information collectée est plus fiable. Incontestablement, la traduction comporte le risque d'introduire un biais d'interprétation des propos des participant-e-s. Une alternative serait de former des acteurs et actrices de la communauté à la démarche et d'inclure ce volet de renforcement des capacités dès la planification de projet. Ainsi, la validité de la démarche, son ancrage communautaire et la durabilité seront renforcés.

Le caractère innovant de la méthode et son contexte d'application : des éléments indispensables à prendre en compte

Dans le contexte du projet, les communautés avaient l'habitude de participer à des activités de suivi et d'évaluation traditionnelles (enquêtes démographiques) effectuées par les intervenant-e-s, et n'avaient jamais été confrontées à une méthode innovante comme la photovoix. Celle-ci met en avant la créativité, la réflexion basée sur le souvenir, et implique que les participant-e-s expriment, à partir d'un exemple concret, leurs émotions et leurs opinions sur les activités décrites sur les photos.

Néanmoins, cette méthode était relativement déstabilisante au départ pour certains participant-e-s. Les facilitateurs et les facilitatrices ont donc accordé une attention particulière à l'explication des objectifs et du déroulement de l'activité. Il était également essentiel de procéder à une écoute active et d'être en mesure de stimuler la réflexion des participant-e-s pour qu'ils et elles aillent au bout de leur pensée et ne se limitent pas à une simple description de la photo choisie.

Par ailleurs, les participant-e-s étant jeunes et n'ayant pas eu l'occasion d'être confrontés au matériel visuel qu'est la photographie, il leur a fallu du temps au début de l'activité pour se familiariser avec les photos, réagir,

échanger, et parfois rire, avant de pouvoir se concentrer à nouveau et répondre à la consigne de l'exercice. C'est notamment pour cela que dans la deuxième communauté, un jeu brise-glace a été organisé avec les jeunes à partir des photos pour les mettre à l'aise et les préparer à l'activité avant de solliciter la leur pleine concentration. La prise en compte du contexte dans lequel est menée l'activité et de l'audience, en l'occurrence peu habituée à des méthodes d'évaluation innovantes et au matériel photographique, est donc essentielle dans la planification de l'activité. Elle nécessite une bonne capacité d'adaptation et d'ajustement en temps réel de la part de l'animateur ou de l'animatrice. Une bonne connaissance de la réalité culturelle et sociale où se réalise l'activité permet de meilleures organisation, planification et participation à l'activité.

L'analyse de l'activité : un processus long et complexe

L'analyse des données issues de la photovoix (prise de notes, enregistrements de l'activité, propos des participant-e-s, photos sélectionnées) a constitué un défi important compte tenu de la longueur du processus et de l'importance d'obtenir des données valides et utilisables rapidement.

Pour être en mesure d'analyser les propos des participant-e-s, la prise de notes a été privilégiée dans la première communauté. Toutefois, cette tâche s'est avérée complexe, en particulier en raison de la barrière de la langue (le preneur de notes ne comprenait pas le nahuatl) et de la rapidité des échanges. Dans le cas de la deuxième communauté, il a donc été décidé d'enregistrer et de filmer l'activité avec l'accord préalable des participant-e-s et de leurs parents afin de pouvoir observer *a posteriori* les images et mieux rendre compte du déroulement et des résultats de l'activité.

Par la suite, et comme décrit précédemment, plusieurs étapes d'analyse ont été menées afin de répondre aux objectifs fixés (recueillir les perceptions des participant-e-s par rapport aux ateliers d'art social et mieux comprendre les effets de l'art social sur l'acquisition de connaissances et d'aptitudes artistiques). Le processus a nécessité du temps, des ressources humaines spécialisées (en recherche et évaluation participative) et disponibles pour réaliser une analyse intégrée et participative prenant en compte la complexité du contexte d'intervention et une bonne connaissance du projet.

Conclusion

La photovoix a permis de répondre à l'objectif d'évaluation qui consistait à mieux comprendre les perceptions des participant-e-s relativement aux ateliers artistiques et pédagogiques centrales aux projets d'art social de One Drop. Participative, ludique et favorisant le dialogue, cette démarche évaluative a été respectueuse des participant-e-s et s'est adaptée au contexte d'implantation tout en inscrivant dans la continuité des activités artistiques.

> Il peut être difficile pour les chercheuses et chercheurs d'observer certains des comportements les plus importants d'une communauté, en particulier lorsque cette communauté se méfie des étrangers, est isolée, mal desservie, opprimée ou opérant sans langage écrit. Dans ces situations, une recherche observationnelle menée par des étrangers peut conduire à une vision biaisée et inexacte de la communauté qui peut alors aboutir à une mauvaise compréhension de ses besoins, de ses atouts et de ses valeurs culturelles (...) [La photovoix] est très utile lorsqu'elle est employée au début d'un projet, ainsi que pendant l'étape d'évaluation et d'analyse pour mesurer le succès et la validité. (Haldenby, T. date non renseignée, traduction libre[5]).

Même si les participant-e-s n'étaient pas les auteurs et autrices des photographies, ils et elles en étaient les acteurs et actrices et à ce titre, ont pu se découvrir dans leur environnement, en action et en collaboration avec leurs pair-e-s. L'image a suscité des réactions spontanées avant de favoriser la réflexion et la discussion. La voix a été donnée aux jeunes participant-e-s à travers la photo et le dessin, des canaux privilégiés pour le partage d'expérience et l'expression de ses perceptions, émotions, et opinions

5. *It can be difficult for researchers to observe some of a community's most important behaviors, particularly when that community is wary of outsiders, isolated, underserved, oppressed or operating without written language. In these situations, observational research conducted by outsiders can result in a skewed and inaccurate view of the community, which can then result in a poor understanding of its needs, assets and cultural values (...) [Photovoice] is most useful when employed at the front-end of a project, as well as during the evaluation and analysis stage to gauge success and validity.* http://designresearchtechniques.com/casestudies/photovoice/ consulté en octobre 2016

souvent non prises en compte par les adultes. La mise en œuvre de cette démarche présente des défis qui ne remettent pas en cause son potentiel d'apprentissage et transformation individuelle et collective.

À la suite de cette expérience, One Drop a opté pour l'utilisation de la photovoix comme intervention en plus de l'appliquer comme démarche évaluative, car elle comporte les avantages des interventions basées sur les arts (Jagiello, 2015; Coholic et al. 2012; Djurichkovic 2011). Ainsi, elle rejoint l'ensemble des activités d'art social (théâtre d'intervention, ateliers artistico-pédagogiques, documentation et média) utilisées dans ses projets d'accès à l'eau, l'hygiène et l'assainissement.

Enfin, cette activité a aussi généré des produits visuels (photos, vidéos, dessins) et permis de fournir des informations plus générales sur la dynamique sociale au sein des communautés ciblées. Les résultats ont été diffusés auprès des partenaires du projet, de la communauté, mais aussi auprès d'acteurs et actrices du secteur eau, hygiène et assainissement.

Références clés

Wang, C. et Burris, M. A. (1997). Photovoice : Concept, methodology, and use for participatory needs assessment. *Health education & behavior*, 24(3), 369-387.

Dans cet article, Wang et Burris, considérées comme des pionnières de la photovoix, dressent un portrait global de la méthodologie. Elles en présentent les origines, les avantages et les limites. Ensuite, les implications pratiques, les éléments facilitants et les contraintes sont analysées à la lumière d'un projet d'évaluation participative des besoins mis en œuvre au sein d'une communauté chinoise.

Fradet, L. (2012). *Guide de la méthode photovoix*. https://sites.google.com/site/rappratique/outils-et-documentation/guide-methode-photovoix

Dans ce guide pratique, Lucie Fradet présente étape par étape la méthode photovoix dans le cadre d'une recherche-action participative. De la préparation à l'évaluation, en passant par la collecte et l'analyse des données, la photovoix est traduite en actions. Le guide vulgarise les concepts théoriques et méthodologiques associés à la photovoix tout en offrant à toute personne souhaitant l'utiliser des « trucs et astuces» pour en faire un succès.

Mackinnon, S. et Ryan, L. (2015). Photovoice: Through the lens – A case study. *Youth Engagement in Health Promotion*. Vol1. No.1

Mackinnon et Ryan démontrent, dans cette étude de cas conduite en Ontario, la pertinence d'utiliser la photovoix comme outil favorisant la discussion avec des jeunes. A travers l'alliance de la photographie et des récits, leurs perceptions sont prises en compte dans la définition de ce qu'est une communauté en santé. La photovoix est également présentée comme une stratégie de promotion de dialogue avec les décideuses et décideurs au sein de la communauté. L'auteur conclut que la photovoix est une méthode de recherche-action participative particulièrement efficace en promotion de la santé, notamment à travers le concept d'empowerment (autonomisation) des communautés.

Références

Catalani, C. et Minkler, M. (2010). Photovoice: A review of the literature in health and public health. *Health education & behavior*, 37(3), 424-451.

Coholic, D., Eys, M. et Lougheed, S. (2012). Investigating the effectiveness of an artsBased and mindfulness-based group program for the improvement of resilience in children in need. *Journal of Children and Family Studies*, 21, 833–844.

Danchin, É. et Sax, C. (2009). *Photolangage sur les sentiments d'insécurité et de sécurité*. Bruxelles. BE: Cultures & Santé asbl.

Delgado, M. (2015). *Urban youth and photovoice: Visual ethnography in action*. Oxford : Oxford University Press.

Desyllas, M. C. (2014). Using photovoice with sex workers: The power of art, agency and resistance. *Qualitative Social Work*, 13, 477-501.

Djurichkovic, A. (2011). *Arts in prisons: A literature review of the philosophies and impacts of visual arts programs for correctional populations*. https://opus.lib.uts.edu.au/bitstream/10453/19836/3/ArtinPrisons_Djurichkovic.pdf

Duffy, L. (2010). Hidden heroines: Lone mothers assessing community health using Photovoice. *Health Promotion Practice*, 11(6), 788–797.

Figueroa, M.E. et Kincaid, D.L. (2010). *Social, Cultural and behavioral*

correlates of household water treatment and storage. http://ccp.jhu.edu/wp-content/uploads/Household-Water-Treatment-and-Storage-2010.pdf

Fradet, L. (2012). *Guide de la méthode photovoix.* https://sites.google.com/site/rappratique/outils-et-documentation/guide-methode-photovoix

Freire, P. (1970). *Pedagogy of the oppressed* (traduit par M. Ramos). New York : Herder & Herder.

Jagiello, J. (2015). A *case study: art-based interventions into management education.* http://pin-c.sdu.dk/assets/art-based-interventions-into-management-education–jolanta-jagiello–16-23–pinc-2015.pdf

Goo-Kuratani, D. L. et Lai, E. (2011). *Photovoice literature review.* http://teamlab.usc.edu/PhotovoiceliteratureReview(FINAL).pdf

Mackinnon, S. et Ryan, L. (2015). Photovoice: Through the lens – A case study. *Youth Engagement in Health Promotion*, 1(1), p-p.

Madrigal, D., Salvatore, A., Casillas, G., Casillas, C., Vera, I., Eskenazi, B. et Minkler, M. (2014). Health in my community: Conducting and evaluating photovoice as a tool to promote environmental health and leadership among Latino/a youth. *Progress in community health partnerships: research, education, and action*, 8(3), 317–329. http://doi.org/10.1353/cpr.2014.0034

Mayfield-Johnson, S., Rachal, J. R. et Butler, J. (2014). When we learn better, we do better: Describing changes in empowerment through photovoice among community health advisors in a breast and cervical cancer health promotion program in Mississippi and Alabama. *Adult Education Quarterly*, 64(2), 91-109.

Moyer-Gusé, E. (2008). Toward a theory of entertainment persuasion: Explaining the persuasive effects of entertainment-education messages. *Communication Theory*, 18(3), 407-425.

Onuekwe, C. E. (2015). *Entertainment-education for HealthBehaviour Change.* Lieu d'édition : Friesen Press.

Palibroda, B., Krieg, B., Murdock, L. et Havelock, J. (2009). A *practical guide to*

photovoice: sharing pictures, telling stories and changing communities. http://www.pwhce.ca/photovoice/pdf/Photovoice_Manual.pdf

Rania, N., Migliorini, L., Rebora, S. et Cardinali, P. (2014). Enhancing critical dialogue about intercultural integration: The Photovoice technique. *International Journal of Intercultural Relations*, 41, 17-31.

Singhal, A. et Rogers, E. M. (1999). *Entertainment-education: A communication strategy for social change*. Mahwah, NJ: Lawrence Erlbaum Associates.

Strack, R. W., Lovelace, K. A., Jordan, T. D. et Holmes, A. P. (2010). Framing photovoice using a social-ecological logic model as a guide. *Health promotion practice*, 11(5), 629-636.

Strack, R. W., Magill, C. et McDonagh, K. (2004). Engaging youth through photovoice. *Health Promotion Practice*, 5(1), 49-58.

Sutton-Brown, C. A. (2014). Photovoice : A methodological guide. *Photography and Culture*, 7(2), 169–186

Vacheret, C. (2011). Le photolangage, une médiation thérapeutique : un bref historique des théories groupales. *Le Carnet PSY*, 141(1), 39-42.

Valiquette-Tessier, S. C., Vandette, M. P. et Gosselin, J. (2015). In her own eyes: Photovoice as an innovative methodology to reach disadvantaged single mothers. *Canadian Journal of Community Mental Health*, 34(1), 1-16.

Wang, C. C. (2006). Youth participation in photovoice as a strategy for community change. *Journal of Community Practice*, 14(1-2), 147-161.

Wang, C. C. et Redwood-Jones, Y. A. (2001). Photovoice ethics : Perspectives from Flint photovoice. *Health education & behavior*, 28(5), 560-572.

Wang, C. et Burris, M. A. (1997). Photovoice : Concept, methodology, and use for participatory needs assessment. *Health education & behavior*, 24(3), 369-387.

Wang, C., Yuan, Y. L. et Feng, M. L. (1996). Photovoice as a tool for participatory evaluation : the community's view of process and impact. *Journal of contemporary health*, 4, 47-49.

White, R., Sasser, D., Bogren, R. et Morgan, J. (2009). Photos can inspire a thousand words: Photolanguage as a qualitative evaluation method. *Journal of Extension*, 47(3), 1-5.

World Health Organization et UNICEF. (2015). *Progress on sanitation and drinking water: 2015 update and MDG assessment.* http://files.unicef.org/publications/files/Progress_on_Sanitation_and_Drinking_Water_2015_Update_.pdf

Résumé / Resumen / Abstract

Utiliser des méthodes évaluatives adaptées à la fois au contexte d'intervention, à l'objet de l'évaluation et aux populations ciblées est essentielle en évaluation de programmes. Toutefois, cela comporte des défis non négligeables dans le cas d'interventions complexes et de communautés marginalisées. Dans ce chapitre, nous présentons une méthode d'évaluation participative utilisant la photo comme support d'expression -la photovoix-, particulièrement adaptée aux jeunes et aux communautés autochtones ciblées par un projet d'accès à l'eau, l'hygiène et l'assainissement au Mexique. Cette méthode qualitative a aussi été choisie pour son caractère créatif et artistique permettant de répondre à l'objectif de l'évaluation : évaluer les perceptions des jeunes par rapport aux activités artistiques auxquelles ils ont participé dans le cadre de la composante Art Social pour le Changement de Comportement ; les encourageant à adopter de meilleures pratiques en lien avec l'eau potable et l'hygiène. Les forces et les faiblesses de cette méthode sont exposées et le processus méthodologique décrit en détail. Les résultats de l'intervention mis en évidence par ce processus sont brièvement présentés et enfin, les leçons tirées de l'application de la photovoix sont décrites afin d'encourager l'utilisation de cette méthode qui s'est avérée particulièrement utile et concluante dans le cas présenté.

***

Para la evaluación de programas, es esencial usar métodos evaluativos que sean adaptados al contexto de la intervención, al objetivo de la evaluación y a las poblaciones meta. Sin embargo, esto representa un desafío para intervenciones complejas y para comunidades marginadas. En este capítulo, presentamos un método de evaluación participativa que utiliza la foto como apoyo para la expresión, llamado fotovoz, el cual está particularmente adaptado a los jóvenes y a las comunidades indígenas de un proyecto de acceso a agua, saneamiento e higiene en México. Se escogió este método también por sus características creativas y artísticas, alineado con el

objetivo de la evaluación : evaluar las percepciones de los jóvenes en cuanto a las actividades artísticas en las cuales participaron en el contexto del componente de Arte Social para el Cambio de Comportamiento del proyecto, el cual promovía buenas prácticas relativas al agua potable y a la higiene. Se identifican las fortalezas y debilidades de este método y se describe el proceso metodológico en detalle. Se presentan brevemente los resultados de la intervención puestos en evidencia por este proceso y finalmente se describen las lecciones aprendidas de la aplicación de la fotovoz, con el fin de promover la utilización de este método que resultó muy útil y exitoso en el caso presentado.

***

In program evaluation, it is essential to use evaluation methods which are adapted to the intervention context, the evaluation objective and the target populations. However, this is quite challenging in the case of complex interventions and of marginalized communities. In this chapter, we are presenting a participatory evaluation method, called photovoice, using photos as a support for expression. This choice was particularly adapted to the youth and the indigenous communities targeted by a project of access to water, sanitation and hygiene in rural Mexico. This qualitative method was also selected for its creative and artistic characteristics, well suited to the evaluation objective : to assess the youth's perceptions regarding the artistic activities they had participated in, as part of the Social Art for Behavior Change component of the project, to promote the adoption of adequate practices related to safe water and hygiene. The method's strengths and weaknesses are outlined and the methodological process is described in detail. Findings from the evaluation are briefly presented, as well as lessons learned from the use of photovoice, in an effort to promote the application of this method which proved to be very useful and successful in the case studied.

***

Lynda Rey est spécialisée dans l'utilisation d'approches innovantes en évaluation de programmes et planification stratégique. Elle réalise des mandats diversifiés en santé publique, développement international et gestion des ONG notamment auprès d'organisations publiques, privées et communautaires, de fondations et institutions académiques à travers le monde. Récemment, elle a dirigé l'unité de suivi, évaluation et gestion des

connaissances à One Drop, une fondation basée à Montréal qui investit dans des projets d'accès à l'eau, l'hygiène et l'assainissement intégrant l'art social pour le changement de comportements. Lynda Rey est titulaire d'un doctorat en Santé publique et d'un diplôme en Analyse et Évaluation des interventions en santé de l'Université de Montréal. Elle a également obtenu deux maîtrises en France; l'une en Sciences politiques et relations internationales de Sciences-Po Aix et l'autre en Coopération internationale et Politiques de développement de la Sorbonne.

Wilfried Affodégon est ingénieur agroéconomiste. Il est également titulaire d'un master en Management. Actuellement, il est candidat au doctorat en Science politique à l'Université Laval, Québec-Canada. Il a travaillé pendant environ une décennie pour les programmes d'aide au développement international en tant que spécialiste en suivi et évaluation. À ce titre, il a utilisé la vidéo participative qui est une variante de l'approche photovoix pour l'évaluation des appuis de la coopération danoise aux communautés agricoles au Bénin. Il est également spécialiste des institutions parlementaires, du management public, de la mesure de la performance et de l'évaluation de l'action publique ainsi que des méthodes de recherche en science politique.

Isabelle Viens est diplômée en arts dramatiques de l'Université du Québec à Montréal. Elle collabore, depuis plus de 25 ans, à des projets de développement international qui l'ont amenée à vivre des expériences riches et humaines en travaillant à partir de l'art comme source de transformation sociale. Depuis les 10 dernières années, elle a fait de l'Amérique Centrale sa maison. Actuellement, elle est gestionnaire d'interventions en art social pour le changement de comportement en Amérique latine pour la Fondation One Drop.

Hind Fathallah est spécialiste en évaluation de projets de développement international. Elle est présentement conseillère en suivi, évaluation et apprentissage auprès de la fondation One Drop à Montréal, où elle combine son expertise en développement international et son expérience en évaluation de projets en art social et en changement de comportements liés à l'eau, l'hygiène et l'assainissement, en Afrique de l'Ouest, en Amérique latine et en Inde. Elle est passionnée par le développement international et les méthodes évaluatives participatives et innovantes. Elle détient une maîtrise en coopération internationale de

Sciences Po Lyon (France) et a travaillé auprès de diverses organisations à but non lucratif en France et au Canada.

Maria José Arauz est spécialiste en évaluation de programmes. Elle détient une maîtrise en santé publique de l'Université de Montréal et possède plus de 7 ans d'expérience en évaluation de programmes et recherche évaluative. Maria a travaillé dans des projets de développement international, de changement de comportement, de santé publique, de santé mentale, de déterminants sociaux de la santé et de développement social et auprès des groupes divers comme des populations rurales, autochtones et immigrantes. Maria favorise les approches participatives, créatives et axées sur l'utilisation des résultats de l'évaluation.

Citation

Lynda Rey, Wilfried Affodégon, Isabelle Viens, Hind Fathallah et Maria-Jose Arauz. (2019). La méthode photovoix. Une intervention auprès de populations marginalisées sur l'accès à l'eau potable, l'hygiène et l'assainissement au Mexique. In *Évaluation des interventions de santé mondiale. Méthodes avancées*. Sous la direction de Valéry Ridde et Christian Dagenais, pp. 83-123. Québec : Éditions science et bien commun et Marseille : IRD Éditions.

5. L'analyse d'une recherche-action

Combinaison d'approches dans le domaine de la santé au Burkina Faso

AKA BONY ROGER SYLVESTRE, VALÉRY RIDDE ET LUDOVIC QUEUILLE

L'analyse d'une recherche-action

Définition de la méthode

Il s'agit d'analyser la démarche de mise en œuvre d'une recherche-action en combinant une analyse externe menée par une personne extérieure à la recherche-action à une analyse interne (analyse réflexive) faite par les acteurs et actrices de la recherche-action.

Forces de la méthode

- L'analyse externe est factuelle et basée sur des critères objectifs
- L'analyse interne (analyse réflexive) permet aux acteurs et actrices de la recherche-action de mener des réflexions pour approfondir l'analyse externe et d'apporter des éléments supplémentaires
- La combinaison des deux types d'analyses permet d'améliorer la qualité de l'analyse par l'utilisation des deux analyses complémentaires et à recourir à une triangulation de méthodes pour collecter des informations complémentaires

Défis de la méthode

- Pouvoir concevoir une grille d'analyse externe capable de comprendre la réalité, de servir de boussole à l'analyse externe
- Être capable d'amener les parties prenantes à s'approprier la démarche d'analyse et de faire la distinction entre « objet de recherche » et « démarche de mise en œuvre d'une recherche-action », en utilisant un vocabulaire adapté afin de produire des connaissances utiles
- Être capable de réunir toutes les parties prenantes pour mener une analyse réflexive, sans se heurter aux contraintes de calendrier et de ressources financières

Depuis 2008, deux districts sanitaires de la région du Sahel au Burkina Faso (Sebba et Dori) bénéficient d'un projet pilote d'exemption du paiement des soins en faveur des enfants de moins de cinq ans et des femmes enceintes et allaitantes, afin d'améliorer l'accès aux soins de ces populations vulnérables. Mais si les soins sont gratuits pour ces patients et patientes, il faut que les dépenses effectuées à leur endroit par les centres de santé soient intégralement remboursées à ces derniers pour ne pas mettre le

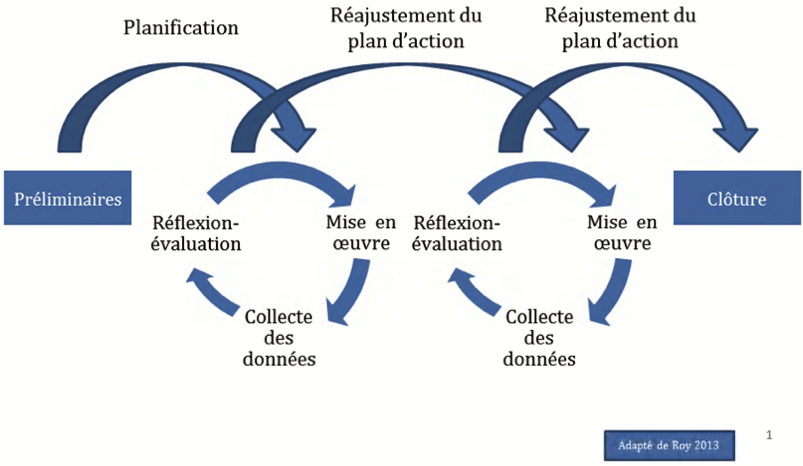
système de santé en faillite. Le remboursement au réel à travers le calcul de tous les frais des prestations réalisées par les Centres de santé présente cependant de nombreux inconvénients : une charge de travail pour le personnel de santé et un coût élevé de production des outils pour l'ONG subventionnaire.

Pour y faire face, une équipe de chercheuses et chercheurs mène depuis 2011, en collaboration avec les parties prenantes de ce projet, une recherche-action dans le district sanitaire de Sebba. Elle vise à expérimenter une innovation, à savoir le « remboursement au forfait » des prestations des centres de santé. Cela consiste à leur octroyer une somme forfaitaire pour leurs prestations qui est multipliée, pour la demande de remboursement, par le nombre de patients et patientes reçus.

Dans ce chapitre, nous analysons la démarche utilisée pour mettre en œuvre cette recherche-action afin d'en tirer des leçons pouvant guider de futurs projets similaires. Il s'agit d'offrir aux lecteurs et lectrices un cadre d'analyse original afin de montrer comment il est possible d'évaluer et analyser un processus de recherche-action en santé mondiale.

La recherche-action est un processus itératif impliquant la collaboration entre chercheuses, chercheurs et praticien-ne-s pour agir ensemble dans un cycle d'activités comportant l'identification du problème, la planification, la mise en œuvre, l'évaluation des solutions et la réflexion (Avison 1999).

Roy et Prévost (2013) a représenté les différentes étapes des cycles expérimentaux de la recherche-action en ajoutant une étape préliminaire et une étape finale au modèle originel de Lewin (1940). Les cycles expérimentaux prennent fin si les parties prenantes ont résolu tous les problèmes identifiés.



Les cycles d'une recherche-action. Source : Roy 2013

La démarche utilisée pour la mise en œuvre de cette recherche-action a regroupé un ensemble d'informations. Le modèle de Roy (2013) nous a permis de la décomposer en plusieurs étapes et cycles pour bien la comprendre avant de nous concentrer sur la démarche utilisée pour mettre en œuvre les différentes étapes. La notion de « démarche » est définie par Lui (dans Boudjaou et Bourassa, 2012) comme « le processus de réalisation d'intentions de recherche et de changement dans lequel les acteurs et actrices possèdent une grande autonomie à la fois sur les fins et les moyens ».

Ce chapitre ne prend pas en compte l'objet de recherche (expérimentation du remboursement au forfait et ses effets) qui fait l'objet d'une autre étude (non publiée). Il se concentre sur l'analyse de la démarche utilisée pour réaliser cette recherche-action dans le cadre d'une collaboration ou d'un partenariat entre chercheuses, chercheurs et parties prenantes (Reason, 2011).

Dans le contexte de ce projet, l'idéal était de réunir toutes les parties en présence (ONG, personnel de santé à divers échelons, représentants de

la population) pour mener une analyse réflexive sur la démarche utilisée pour mettre en œuvre la recherche-action et en tirer les leçons. Cependant, l'utilisation seule d'une telle analyse se heurte à plusieurs défis. La plupart des parties prenantes possèdent trop peu de connaissances scientifiques sur la démarche de mise en œuvre d'une recherche-action pour mener à bien cette analyse. Une analyse externe a été jugée nécessaire en se basant sur des critères factuels pour bien la mener en toute indépendance. Cependant, d'après les approches utilisées dans le domaine de l'évaluation, mener seulement une analyse externe pourrait également entraîner de nombreux défis, tels que la difficile utilisation des résultats. À ce sujet, Bonami (2005) précise:

> De nombreux rapports d'évaluation ou de conseils meublent les armoires de décideuses et décideurs avant de disparaître aux archives ou ailleurs. Il est moins banal qu'il n'y paraît de rappeler que l'aboutissement attendu des démarches d'évaluation tant interne qu'externe réside dans les prises de décision qui leur font suite et qui renforcent, confirment, infléchissent, modifient les pratiques pédagogiques et de gestion en cours dans les établissements scolaires.

Au regard de ce qui précède, notre analyse s'est déroulée en trois étapes : i) synthèse des connaissances sur la démarche de réalisation d'une recherche-action et identification des meilleures pratiques pour les réaliser, ii) analyse externe de la démarche de la RA à l'aide de la grille élaborée de meilleures pratiques et iii) analyse réflexive des parties prenantes sur la démarche utilisée pour conduire la RA.

Contexte et programme évalué

Description du contexte où se déroule le programme

Depuis 2008, les districts sanitaires de Sebba et de Dori bénéficient d'un projet pilote d'exemption du paiement des soins en faveur des enfants de moins de cinq ans et des femmes enceintes et allaitantes, deux groupes qui ont un faible accès aux soins de santé. Les centres de santé (CS) ont la responsabilité de leur fournir gratuitement des soins de santé (consultations,

accouchements, hospitalisations, soins médicaux et infirmiers, médicaments, etc.). En contrepartie, l'ONG allemande Help, bénéficiant d'une subvention de l'Office Humanitaire de la Communauté Européenne (ECHO) de l'Union européenne, s'est engagée comme tiers payeur. Des difficultés inattendues dans la gestion de ce projet sont vite apparues. En effet, tous les frais afférents à la prise en charge par les centres de santé de ces patients et patientes bénéficiaires sont calculés de façon manuelle dans les moindres détails avant de faire l'objet d'une demande de remboursement auprès de l'ONG. Au début du projet, des documents administratifs comportant plusieurs pages étaient édités et faisaient l'objet de remplissage par le personnel de santé lors du passage des enfants dans les centres de santé. Un effet attendu de cette gratuité fut l'augmentation du taux de fréquentation des centres de santé par la population cible. L'augmentation de la quantité de documents à remplir a alors entraîné, de manière non attendue, une augmentation de la charge du travail du personnel de santé. Les coûts liés à la production des documents devenaient également exorbitants pour l'ONG, ce qui rallongeait les délais de remboursement des centres de santé. Cette situation a fait l'objet de plusieurs plaintes exprimées par le personnel de santé, les membres des comités de gestion (COGES) des centres de santé et des acteurs et actrices de l'ONG. Devant la récurrence de ces plaintes, plusieurs solutions ont été tentées, sans grand succès, notamment la réduction du nombre de documents à remplir et l'installation dans les centres de santé de grosses cantines pour servir à l'archivage des documents (plus de place de stockage dans les bureaux).

Devant la persistance des effets inattendus du remboursement au réel et des plaintes exprimées, les parties prenantes du projet, avec l'aide d'une équipe de recherche de l'Université de Montréal, ont convenu de conduire une recherche-action à partir de 2011 pour expérimenter une innovation à savoir le « remboursement au forfait ».

Description du programme

La recherche-action a concerné seulement la prise en charge des enfants. L'action proposée consistait à verser une somme forfaitaire de 1300FCFA (environ 2,7 $US) par enfant et par mois au centre de santé. Ce montant a été déterminé sur la base des coûts moyens mensuels de prise en charge des enfants au cours de la période allant de janvier 2009 à juin 2011

dans les six premiers centres de santé ayant expérimenté le remboursement au forfait. Sur la base de cette somme, les centres de santé devaient compter et multiplier le nombre d'enfants soignés par la somme forfaitaire puis faire la demande de remboursement auprès de l'ONG. Les agents et agentes de santé devaient utiliser les documents ou outils habituels de collecte de données du ministère de la Santé pour enregistrer les jeunes patients et patientes et ainsi contribuer à la réduction de leur charge de travail et des coûts de production des outils. L'ONG devait faire le transfert de la gestion des remboursements au district sanitaire de Sebba et en assurer le contrôle régulier. Quant à lui, le volet recherche a porté sur i) l'étude de la faisabilité technique et opérationnelle du remboursement au forfait, ii) le calcul de l'évolution des coûts moyens mensuels de prise en charge des enfants et iii) l'évaluation de la charge du travail.

Trois phases ont été planifiées dans cette recherche-action (voir Tableau I). Pour la première phase, six centres de santé sur 13 ont été recrutés. Pour les mettre en confiance, il leur était précisé qu'il n'y avait aucun risque de perte financière. Les centres de santé étaient remboursés sur la base de la somme forfaitaire de 1300 FCFA et du nombre de patients et patientes reçus. Au terme de cette phase, le calcul de tous les coûts réels des prestations des centres de santé devait permettre d'atteindre un équilibre financier. Si les centres de santé avaient plus dépensé par rapport aux sommes perçues au cours du remboursement, l'ONG s'engageait à leur rembourser la différence. Si par contre, les centres de santé avaient perçu plus de remboursement que leurs dépenses réelles, ils s'engageaient à rembourser la différence à l'ONG. Après 11 mois d'expérimentation, les sept autres centres de santé ont été introduits à la deuxième phase de la recherche-action « sans risque financier ». Par contre, les six premiers centres de santé et l'ONG Help avaient pris, à ce moment, « un risque financier », à savoir plus de remboursements en cas de pertes ou de gains financiers. Cette deuxième phase a duré 12 mois. À la troisième phase, les 13 centres de santé avaient pris un risque financier. La durée de cette phase a aussi été de 12 mois. Le district sanitaire de Dori a été choisi comme un district de contrôle durant toute la recherche-action.

Tableau 1. Devis de l'expérimentation du remboursement au forfait (Source : Guide de mise en œuvre de la recherche-action)

Recherche-action	**Phase 1**	**Phase 2**	**Phase 3**
Durée des phases	11 mois	12 mois	12 mois
Groupe 1 (6 CS)	Remboursement sans risque financier	Remboursement avec risque financier	Remboursement avec risque financier
Groupe 2 (7 CS)	Remboursement au réel	Remboursement sans risque financier	Remboursement avec risque financier
District sanitaire de Dori	Remboursement au réel	Remboursement au réel	Remboursement au réel

Méthode d'analyse

L'analyse de la démarche utilisée pour la mise en œuvre de la recherche-action s'est déroulée en trois étapes.

1. Élaboration d'une grille de « meilleures pratiques »

La première étape a consisté en l'élaboration d'une grille de « meilleures pratiques » devant servir à l'analyse de la démarche de la recherche-action. Selon Patton (2001), les « meilleures pratiques basées sur des principes pour guider l'action sont utiles alors que celles qui sont normatives et spécifiques sont difficiles à appliquer dans divers contextes » (traduction libre). Nous avons considéré une « meilleure pratique » comme un savoir-faire, une liste de tâches permettant de réaliser efficacement les différentes étapes des cycles de la recherche-action. Partant des éléments fondamentaux de la définition de la recherche-action, nous avons, à partir de la littérature (scientifique et grise), fait une synthèse de pratiques que nous qualifions de « meilleures pratiques », en plus de celles déjà identifiées par de Roy *et al.* (2013). Les critères de choix de ces pratiques résident dans le fait que ces dernières étaient décrites de façon récurrente par divers auteurs et autrices, tout en démontrant la force de leur preuve.

2. Analyse externe de la démarche de la RA

La deuxième étape consistait à faire une analyse externe de la démarche de cette recherche-action. Nous nous sommes servis de notre grille de « meilleures pratiques » pour faire i) une revue des documents de cette recherche-action, ii) des observations sur le terrain, notamment le déroulement des rencontres de suivi de la recherche-action et des activités dans les centres de santé et iii) des entrevues individuelles avec certains informateurs clés (seulement des hommes) à l'aide d'un guide d'entrevue semi-structurée pour compléter les informations non contenues dans les documents. Avec le coordonnateur de cette recherche-action, nous avons recruté une variété d'informateurs au sein des différentes parties prenantes, choisis selon leur ancienneté, la qualité de leur participation à la recherche-action et la période d'entrée des centres de santé dans les différentes phases de la recherche-action. Notre échantillon se répartissait ainsi :

- Équipe de recherche : le chercheur principal et le coordonnateur de la recherche-action
- ONG Help : le contrôleur financier de la subvention et son adjoint, le coordonnateur médical et un ancien chef du projet
- Direction régionale de la santé du Sahel : le pharmacien
- Équipe-cadre du district de Sebba: le gestionnaire
- Les Centres de santé (CS) : 4 infirmiers-chefs de poste, 4 agents de santé (infirmiers) et 3 gérants de dépôt de médicaments
- 3 Présidents de COGES

Les données ont été collectées à l'aide d'une enregistreuse numérique (dictaphone). Par la suite, elles ont été transcrites et saisies dans le logiciel QDA Miner version 4.1.6. Les résultats de cette deuxième étape nous ont servi à préparer la troisième étape consacrée à l'analyse réflexive des parties prenantes.

3. Analyse réflexive de la démarche de la recherche-action par les parties prenantes

À cette dernière étape, les différentes parties prenantes ont mené une analyse réflexive sur la démarche utilisée pour mettre en œuvre la

recherche-action. Nous avons organisé un atelier de deux jours réunissant, hormis les représentants de la population, toutes les parties prenantes (n = 19). À l'entame de cet atelier, une présentation orale a été faite. Elle a porté sur la définition, les cycles expérimentaux et les étapes d'une recherche-action ainsi que sur les objectifs de l'atelier. Elle a été suivie d'un exercice afin d'amener les parties prenantes à faire la distinction entre l'objet de la recherche-action (« le forfait ») et la démarche utilisée pour mettre en œuvre la recherche-action. Trois activités ont été organisées lors de l'atelier.

La première activité a réuni les parties prenantes en trois sous-groupes homogènes afin de réfléchir aux éléments qu'elles avaient le plus et le moins appréciés dans la démarche et de formuler des recommandations pour de futurs projets de recherche-action. Une restitution des travaux a ensuite été faite, suivie de discussions en plénière.

La deuxième activité a consisté à administrer un questionnaire quantitatif aux parties prenantes afin de recueillir leur degré d'accord par rapport à l'application des « meilleures pratiques » établies (89% de répondants).

La troisième activité a consisté à faire une restitution des résultats préliminaires de l'analyse externe aux parties prenantes, suivie de discussions.

Au terme de cette démarche, nous présentons une synthèse de l'apport de cette méthode.

Résultats de l'analyse

Résultats de la revue documentaire

Nous présentons les résultats obtenus dans une série de tableaux proposant, pour chaque étape du cycle d'une recherche-action, une grille d'analyse des « meilleures pratiques ».

Tableau 2a : Étape préliminaire de la recherche-action

Meilleures pratiques	**Description et preuve de la pratique**
Réaliser une analyse situationnelle	Elle retrace les préoccupations exprimées par les parties prenantes pour justifier la recherche-action. Sa restitution en public aux parties prenantes et les discussions font naitre1) l'intention d'actions (besoin de faire) et 2) l'intention de recherche (besoin de savoir) (Demange 2012; Faure, Hocdé et Chia 2011; Faure et Chia 2011; Faure 2007; Roy 2013; Tousignant 2012; Waterman 2007).
Identifier les parties prenantes	Cette pratique vise à avoir une diversité de points de vue dans la recherche-action et une expression des différents intérêts. Le choix doit être 1) étendu à divers acteurs et actrices et 2) représentatif des acteurs et actrices concernés par la problématique Les critères de choix sont divers (intérêt, fonction, capacité de négocier ou travailler ensemble). La présence des décideuses et décideurs et de leaders d'opinion est fondamentale et facilite la vulgarisation des résultats (Champagne 2007; Demange 2012; Detardo-Bora 2004; Dickens 1999; Faure, Hocdé et Chia 2011; Gauthier 2010; McVicar 2012; Mikolasek, Chia, Pouomogne, Tabi 2009; Roy 2013).
Mettre en place les instances de gouvernance de la recherche-action	Ces instances visent une participation active des parties prenantes. Elles doivent être démocratiques tout en précisant : 1) le rôle de chaque partie prenante et 2) les organes de gestion (comité de pilotage, équipe de recherche...). Le comité de pilotage doit être représentatif des parties prenantes (Casabiaca 1997; Champagne 2007; Faure, Hocdé et Chia 2011; Vespiren 2012).
Formaliser l'engagement réciproque de toutes les parties prenantes	Mettre en place un cadre de partenariat équitable pour établir la confiance. Le contenu du cadre de partenariat doit être transparent et porter sur la gestion de la propriété intellectuelle et des conflits, la confidentialité, les mécanismes de prise de décision, etc. Il est matérialisé par la signature du cadre de partenariat (Casabiaca 1997; Faure, Hocdé et Chia 2011; Le May 2001; Mikolasek, Chia, Pouomogne et Tabi 2009; Roy 2013).

Définir le rôle du chercheur ou de la chercheuse dès le départ	Le chercheur ou la chercheuse a trois responsabilités scientifiques selon Bourassa (2012): • structuration de la recherche-action • accompagnement (animation, médiation, formation, conseils) • production des connaissances (stimule la réflexion, l'analyse des problèmes et des solutions et le passage à l'action) Cette position non partisane permet aux chercheuses et chercheurs d'être des co-acteurs et co-actrices et aux parties prenantes d'être des co-chercheuses et co-chercheurs. Cela permet d'établir la confiance entre les parties prenantes en permettant d'éviter les conflits d'intérêts et conduit à plus d'objectivité (Le May 2001; Mikolasek, Chia, Pouomogne et Tabi 2009; Roy 2013).

Tableau 2b : Planification de la recherche-action

Meilleures pratiques identifiées	Description et preuve de la pratique
Identifier l'objectif de la recherche-action	Formulé sur la base de l'analyse de la situation, cet objectif doit être prioritaire pour le chercheur, la chercheuse et les autres parties prenantes pour susciter leur engagement. Il faut identifier déjà les défis à discuter (Casabiaca 1997; D'Alonzo 2010; Desclaux 2010; Faure, Hocdé et Chia 2011; Gauthier 2010; Roy 2013; Zachariah 2009; Zachariah 2010).
Analyser les enjeux liés à l'objet et à l'implantation de la recherche-action	Selon Lui *et al.* (1997), la présentation de l'objectif de recherche-action aux parties prenantes suivie de discussions permet d'analyser les enjeux liés à l'objet de recherche (faisabilité, financement, participation des parties...), d'obtenir leur accord afin d'établir la confiance avant le démarrage. Pour Le May *et al.* (2001), l'analyse des enjeux permet de sécuriser la collaboration en vue de mettre les participant-e-s en confiance.
Identifier les questions de recherche	Formulées par les chercheuses et chercheurs à partir des objectifs de la recherche-action, les questions de recherche doivent être ouvertes pour produire de façon inductive les connaissances scientifiques. Elles doivent être priorisées selon la logique des entrées et sorties. Elles doivent être traitables au cours d'un cycle de la recherche-action (Casabiaca 1997; Chevalier 2013 ; Faure, Hocdé et Chia 2011; Mikolasek, Chia, Pouomogne et Tabi 2009; Vespiren 2012).
Proposer des solutions adaptées et réalistes	Identifier des scénarios d'évolution des préoccupations exprimées dans l'analyse situationnelle (Faure 2007). Trois caractéristiques fondamentales des solutions pour maintenir la confiance et l'intérêt des parties prenantes sont : 1) innovation, 2) adaptation ou intégration au contexte pour faciliter leur mise en œuvre et 3) fondement sur des preuves ou expériences (Casabiaca 1997; Faure, Hocdé et Chia 2011; Waterman 2007)

Planifier les activités à réaliser	Le protocole de recherche doit être réaliste, souple, non fermé, progressif selon l'évolution de la recherche-action (Roy, 2013). Il comprend les éléments suivants (D'Alonzo 2010; Demange 2012; Faure, Hocdé et Chia 2011; Mshelia 2013; OMS 2008; Waterman 2007) : • devis d'études adapté (type de recherche) • choix des cibles • représentativité des sites d'expérimentation pour le transfert • méthodes de collecte des données adaptées • questions éthiques • identification des indicateurs de mesure à collecter • chronogramme des activités Demange (2012) a identifié deux principes pour son élaboration: 1) élaboration par un groupe restreint (chercheur ou chercheuse, facilitateur ou facilitatrice de recherche, un-e intervenant-e de terrain) et 2) présentation au comité de pilotage pour discussion

Tableau 2c : Mise en place de la recherche-action

Meilleures pratiques identifiées	Description et preuve de la pratique
Élaborer des guides de mise en œuvre des activités	Les activités menées dans le cadre de la recherche-action étant expérimentales, elles doivent être clairement définies, décrites et standardisées dans des guides pour faciliter la mise en œuvre sur tous les sites. Il faut responsabiliser les parties prenantes (Champagne 2007; Demange 2012; Faure, Hocdé et Chia 2011; Faure 2007; Peters et Adam, 2013).
Organiser des séances de travail sur le terrain	Organisation de séances de travail sur le lieu de travail des intervenant-e-s de terrain pour passer en revue les activités menées, le remplissage des outils de collecte de données, détecter d'éventuelles difficultés et y apporter des solutions et renforcer les capacités des acteurs et actrices de terrain au besoin. Le suivi peut aussi se faire à travers le courriel, le téléphone (Champagne 2007; D'Alonzo 2010; Demange 2012; Detardo-Bora 2004; Vespiren 2012).

Tableau 2d : Collecte des données sur la recherche-action

Meilleures pratiques identifiées	Description et preuve de la pratique
Déterminer les méthodes de collecte des données	Selon Dickens *et al.* (1999), les actions dans la recherche-action étant expérimentales, il faut collecter les données utiles et de façon approfondie pour générer les connaissances (Dickens *et al.* 1999; Faure 2007; Mikolasek, Chia, Pouomogne et Tabi 2009). La triangulation (méthodes, données, temporelle et chercheuses et chercheurs) augmente la rigueur scientifique de la recherche-action (Champagne 2007). Il s'agit des entrevues individuelles, de groupes de discussion, d'observations, de revues documentaires (Gauthier 2010; McVicar 2012; Roy 2013). La rétro-information des données stimule la collecte des données (Mshelia 2013).
Élaborer les outils de collecte standardisés	Les outils de collecte sont standardisés en adaptant leur contenu et leur forme (Demange 2012). Cela permet l'exploitation scientifique des données tout en limitant les biais, de comparer les données entre divers sites d'expérimentation. Il peut s'agir de guides d'observation, d'entretien, de questionnaires, de fiche de suivi, de journaux de bord, etc. Il faut faire un pré-test pour évaluer le contenu. L'utilisation des outils existants peut faciliter la collecte des données.
Assurer la traçabilité de la transmission des données en responsabilisant les acteurs et actrices	Toutes les données à collecter sont compilées et transmises selon un circuit déterminé en mentionnant les dates de transmission. Cela permet de retrouver facilement les données égarées. L'archivage des données constitue un atout qui permet de revenir à tout moment sur les données. Cette bonne pratique a été observée sur le terrain.

Tableau 2e : Réflexion-Évaluation de la recherche-action

Meilleures pratiques identifiées	Description et preuve de la pratique
Organiser des rencontres de suivi périodiques	• Organiser des rencontres périodiques dans le cadre du comité de pilotage pour faire le suivi et le bilan de la recherche-action • Avoir une diversité d'opinions dans la validation et l'interprétation des résultats (Toussaint 2012)
Valider ensemble les résultats	• Partager les résultats obtenus après analyse du comité scientifique au comité de pilotage • Engager la discussion entre les parties prenantes pour interpréter les résultats • Critères de validation des résultats selon Chevalier (2013) • Fiabilité (voir triangulation des données) • Consensus (validité de signifiance, sens donné par les acteurs et actrices)
Faire le bilan de chaque cycle du processus	• Discuter et trouver ensemble des solutions aux difficultés rencontrées par les acteurs et actrices de terrain • Redéfinir les questions de recherche innovantes qui proviennent des acteurs et actrices du terrain pour un nouveau cycle • Critères pour mettre fin à un cycle: satisfaction des acteurs et actrices par rapport à la résolution du problème (Roy 2013)
Formaliser les connaissances	• Rédaction de mémos sous forme de journal de bord • Critères de validation des connaissances selon Casabiaca et Roy (Casabiaca 1997; Roy 2013) • Niveau de résolution des problèmes identifiés au départ • Critères de reproduction des connaissances: la transférabilité, la pertinence, la fécondité, l'opérationnalité, les apprentissages.

Une brève description des résultats de l'analyse externe

Étape préliminaire

Les plaintes exprimées par les acteurs et actrices lors de l'exécution du projet d'exemption du paiement des soins étaient les éléments essentiels pour justifier cette recherche-action. En raison de leur persistance (plusieurs solutions infructueuses) et de leur pertinence, elles ont créé une forte intention d'actions (besoin d'agir) parmi toutes les parties prenantes du projet, de même qu'une intention de recherche (besoin de savoir). Même si la recherche d'une solution était prioritaire pour les parties prenantes, l'idée de faire une recherche-action, avancée par l'équipe de recherche, n'a d'abord pas été approuvée par les autres acteurs et actrices du projet, car il y avait trop d'enjeux financiers. Devant ces réticences, les chercheuses et chercheurs ont usé de patience pendant six mois afin d'identifier progressivement les parties prenantes représentatives des différents acteurs et actrices affectés par la problématique au niveau local. Sur le plan national, certains décideuses et décideurs n'ont pas été impliqués, notamment les Programmes Nationaux de Santé Infantile et de Santé Maternelle et la Direction de la Recherche du Ministère de la Santé.

Étape de la planification

Les solutions proposées étaient souples, progressives et adaptées au contexte pour faciliter leur mise en œuvre. Plusieurs étapes ont été nécessaires pour mettre en confiance les différentes parties prenantes, notamment les prestataires de soins et les membres des comités de gestion des structures sanitaires. Ces étapes ont permis de représenter les différents cycles d'action expérimentaux de la recherche-action.

Étape réflexion-évaluation

La projection des résultats de la recherche au cours des rencontres de suivi a permis aux parties prenantes de voir le bien fondé de la recherche-

action et de partager les connaissances. Cependant, l'absence de règles de fonctionnement n'a pas permis de canaliser les débordements dans les propos des parties prenantes tels que les rivalités affichées et les tendances à la stigmatisation des acteurs et actrices qui n'avaient pas eu de bons résultats au cours des rencontres. L'équipe de recherche engagea des discussions ouvertes avec les acteurs et actrices pour interpréter les résultats.

Une brève description des résultats de l'analyse réflexive des parties prenantes

Les parties prenantes ont le plus apprécié leur implication dans la démarche de mise en œuvre de la recherche-action, car chacune a ses spécificités et le fait de toutes les impliquer a permis de réduire les obstacles lors de la mise en œuvre. La signature des conventions a créé la confiance et établi leur engagement. Le choix des groupes de comparaison leur a permis de mieux observer les effets de la recherche-action dans le temps et de dissiper leurs inquiétudes. Le suivi et l'évaluation ont permis un partage d'expériences et de niveau d'information et de prendre ensemble des décisions. De façon unanime, toutes les parties prenantes ont fortement apprécié la mise en œuvre de toutes les meilleures pratiques identifiées à l'étape de réflexion-évaluation.

Cependant, elles ont moins apprécié la non-maîtrise de la durée de la recherche-action qui était liée au financement extérieur, ce qui a induit une multiplicité de conventions. Le retard de la signature des conventions dues aux lenteurs administratives a été parfois source de démotivation. Elles ont également notifié des insuffisances dans la collecte des données, notamment des outils mal renseignés ou égarés. Le retard dans la transmission des termes de références des rencontres de suivi et la courte durée des rencontres ont entraîné une analyse partielle et tardive des résultats.

Analyse réflexive du processus d'analyse

Au terme de notre travail, nous avons conduit une analyse réflexive sur notre démarche et notre processus d'analyse.

Leçons apprises du processus évaluatif

À l'issue du processus d'analyse, les leçons apprises se résument comme suit :

- l'analyse situationnelle au début a créé une intention d'actions et de recherches parmi les parties prenantes
- l'identification et l'implication de toutes les parties prenantes dès le début ont légitimé le processus et ont donné toutes les chances de succès à la recherche-action
- la discussion des enjeux et la signature de conventions ont marqué la confiance et l'engagement des différentes parties en présence durant le processus
- les solutions proposées étaient souples, progressives, innovantes et adaptées au contexte
- l'élaboration d'un guide de mise en œuvre, des visites de terrain et les outils de collecte des données ont facilité les actions expérimentales et la recherche
- l'organisation de rencontres de suivi régulières, démocratiques, encadrées par des règles de fonctionnement a stimulé la réflexion et l'évaluation des résultats par consensus
- la faible implication de certains acteurs et actrices du ministère de la Santé pourrait constituer un défi dans la mise en œuvre des recommandations de cette recherche
- il existe des difficultés dans la collecte et l'organisation des rencontres de suivi de la recherche-action.

Forces et faiblesses de notre processus d'analyse

L'élaboration d'une grille d'analyse des « meilleures pratiques » avec la description de leurs caractéristiques ainsi que la force de leur preuve a permis de les appliquer, de les étudier et de faire une analyse externe factuelle basée sur des critères objectifs à travers une triangulation de méthodes. Les pratiques identifiées proviennent de divers auteurs et autrices et ont été expérimentées dans différents contextes (pays industrialisés et pays en développement) et divers domaines (santé, éducation, agriculture, social) permettant de voir leur adaptation.

La combinaison des deux types d'analyses (externe et interne) offre de nombreux avantages. Elle a permis de structurer la démarche d'analyse en deux étapes complémentaires tout en utilisant une triangulation de méthodes pour collecter des informations complémentaires.

Au cours de l'analyse externe, l'étude de la documentation de la recherche-action n'ayant pas permis d'avoir toutes les informations relatives à la grille des « meilleures pratiques », une triangulation de méthodes (observations, entrevues) a été indispensable pour recueillir des données complémentaires. À ce sujet, les observations ont permis d'observer les enjeux de pouvoir entre les parties prenantes au cours des rencontres de suivi. Cette analyse externe a été utile, car lors de l'analyse réflexive, aucun sous-groupe n'a fait référence aux enjeux de pouvoir, dont la gestion reste déterminante pour la réussite de la recherche-action (Bernier 2014). Cependant, elle n'a pas été suffisante pour recueillir toutes les informations nécessaires. Sans l'analyse réflexive, certaines réalités n'auraient pas pu être décelées par un regard externe. Cette analyse réflexive a permis aux parties prenantes de mener des réflexions approfondies sur certains éléments déjà identifiés lors de l'analyse externe et d'apporter des éléments supplémentaires. À ce sujet, les parties prenantes ont exprimé des préoccupations détaillées relatives à l'organisation pratique de la recherche-action, telles que le retard accusé dans la mise en place du financement, dans l'organisation des rencontres de suivi et dans la signature de conventions, ce qui a parfois été source de démotivation. Les valeurs culturelles des parties en présence les ont certainement obligées à un devoir de réserve face à cette question, surtout en présence des personnes venues d'horizons divers. Cependant, de façon anonyme, une personne s'est exprimée sur le sujet en écrivant en ces termes :

> il faut améliorer la motivation des participant-e-s lors des rencontres de recherche-action parce que ce n'est pas un travail comme les autres, car nous fournissons beaucoup d'effort intellectuel (moins de 15 euros/jour comme jetons de présence lors des rencontres de suivi). Il faut continuer la recherche-action et améliorer les conditions.

Au regard de ce qui précède, nous constatons que la combinaison des deux analyses a amélioré la qualité de l'étude, comme le mentionnent Baron *et al.* :

> Tout d'abord, la participation des groupes concernés par un programme public à son processus d'évaluation permet une meilleure qualité du jugement de valeur qui sera porté sur ce programme, car l'évaluateur aura accès à davantage d'informations au travers de ses échanges avec les participant-e-s et le jugement sera construit à partir d'une multiplicité d'opinions informées (Baron et Monnier, 2003).

Malgré les forces d'une telle analyse combinée, il convient de relever certaines limites. Le fait d'avoir assemblé des « meilleures pratiques » énoncées par divers auteurs et autrices dans le modèle de Roy et Prévost (2013) peut présenter des limites. Certainement, notre recension des écrits n'a pas été complète ni pris en compte tous les articles scientifiques les plus pertinents, ce qui a donné lieu à trois versions de l'outil lors de sa conception. L'absence de consensus concernant la définition de la recherche-action dans les écrits scientifiques a affecté le positionnement des « meilleures pratiques » dans les différentes étapes. Dans la vie, toutes les étapes du cycle de la recherche-action ne sont pas toujours distinctes, notamment les étapes préliminaires, de planification et de réflexion-évaluation.

Principaux enjeux liés à la conduite de l'analyse

Le premier enjeu réside dans la capacité de concevoir une grille d'analyse externe capable de comprendre la réalité, de servir de boussole à l'analyse externe. Dans notre cas, l'absence de consensus sur la définition d'une recherche-action (Gauthier 2010) et des « meilleures pratiques » fut une difficulté. Le deuxième enjeu concerne la capacité de réunir toutes les parties prenantes pour mener une analyse réflexive, sans être confrontés aux contraintes de calendrier et de ressources financières (Baron et Monnier 2003). Dans notre cas, l'atelier a été reporté à deux reprises en raison de conflits de calendrier. Le budget initial de la recherche-action n'ayant pas prévu une telle analyse, il a fallu faire un plaidoyer pour organiser l'atelier.

Le troisième enjeu traite de la capacité d'amener les parties prenantes à s'approprier la démarche d'analyse et à faire la distinction entre un objet de recherche et une démarche de mise en œuvre d'une recherche-action, en utilisant un vocabulaire adapté afin de produire des connaissances utiles.

Dans notre cas, pour éviter d'influencer l'analyse des parties prenantes, les séances ont été dirigées par certaines d'entre elles. Cependant, étant donné qu'elles étaient peu impliquées dans les préparatifs, pour plusieurs raisons, leur participation a été peu remarquable.

Comment procéder la prochaine fois?

Si c'était à refaire, il serait judicieux de bien identifier les mots clés pour faciliter la recension des écrits et établir des critères de sélection plus rigoureux des articles. S'agissant de notre démarche de travail, la qualité de notre analyse pourrait être biaisée par les relations fortes qui existent entre l'équipe de recherche-action et les parties prenantes depuis un certain nombre d'années dans cette zone.

Au regard des limites, il conviendrait de diffuser cet outil d'analyse, de le tester davantage dans divers contextes et de le réviser si possible. L'atelier organisé n'était pas suffisamment préparé avec les co-animateurs et co-animatrices externes, notamment avec les représentants de l'ONG et le responsable du District sanitaire. On aurait pu avoir plus d'informations au niveau de l'analyse réflexive pour générer plus de connaissances. Les consignes données aux différents groupes de travail n'étaient pas suffisamment claires au début. Si c'était à refaire, il aurait fallu leur donner l'occasion de bien clarifier les consignes afin qu'ils soient au même niveau d'informations. En raison de la grande quantité de documents non archivés, nous n'avons certainement pas consulté tous les documents qui auraient pu éclairer davantage notre lanterne.

Conclusion

La mise en œuvre d'une recherche-action comporte différents éléments qui peuvent constituer des facteurs limitant l'atteinte des objectifs s'ils ne sont pas correctement identifiés et bien gérés.

L'analyse de la démarche utilisée pour mettre en œuvre une recherche-action est utile pour guider les futurs projets de recherche-action, mais elle reste confrontée à de nombreux défis. Ce travail a permis de structurer la démarche d'analyse en trois étapes complémentaires. L'élaboration d'une grille de « meilleures pratiques » a constitué une boussole. La combinaison

de l'analyse externe avec l'analyse réflexive de parties prenantes offre de nombreux avantages en termes de complémentarité des données à collecter et de la qualité de l'analyse, en recueillant le point de vue des parties de façon plus approfondie à travers une triangulation de méthodes. Pour réaliser une telle démarche d'analyse, il convient de bien sélectionner la documentation afin de constituer une grille d'analyse externe basée sur des éléments factuels. Pour faciliter la réflexion des parties prenantes, il faut les impliquer depuis le début du processus afin de bien préparer l'atelier de réflexion et utiliser un vocabulaire adapté.

Références clés

Gauthier, B. (2010). La recherche-action. B. Gauthier (dir.), *Recherche sociale : de la problématique à la collecte des données* (p. 517-533). Québec : Presses de l'Université du Québec.

Un chapitre ancien et excellent car un des premiers à aborder la recherche-action comme une démarche scientifique en sciences sociales en langue française.

Reason, P. et Bradbury, H. (dir.). (2011). *Handbook of action research: Participative inquiry and practice*. Londres : Sage Publications.

La bible pour les personnes qui souhaitent comprendre les détails des fondements épistémologiques, théoriques et conceptuels de la recherche-action. L'ouvrage regorge aussi de multiples exemples montrant comment les recherches-actions sont mises en œuvre dans plusieurs contextes et domaines, y compris dans des pays du Sud. Une section concerne aussi les compétences et la formation des acteurs et actrices de la recherche-action.

Roy, M. et Prévost, P. (2013). La recherche-action : origines, caractéristiques et implications de son utilisation dans les sciences de la gestion. *Recherches qualitatives*, 32(2), 129-151.

Un article relativement récent qui permet de mettre au jour les différentes étapes d'une recherche-action et qui a notamment servi de base au présent chapitre.

Références

Avison, D., Lau, F., Myers, M. et Nielsen, P.A. (1999). Action Research: To make academic research relevant, researchers should try out their theories with practitioners in real situations and real organizations. *Communications of the ACM*, 42(1), 94-97.

Baron, G. et Monnier, E. (2003). Une approche pluraliste et participative : coproduire l'évaluation avec la société civile. *Informations sociales*, 110, 120-129.

Baum, F., McDougall, C., Danielle, S. (2006). Participatory action research. *J Epidemiol Community Heath*, 60, 854-857.

Bernier, J. (2014). La recherche partenariale comme espace de soutien à l'innovation. *Global Health Promotion* 2014 21: 58, 21(Supp.1), 58-63.

Boudjaou, M. et Bourassa, B. (2012). Introduction. In Bourassa, B and Boudjaoui, M. (dir.). *Recherches collaboratives en sciences humaines et sociales. Enjeux, modalités et limites*. Québec : Presses de l'Université Laval.

Bourassa, B., Leclerc, C. et Fournier, G. (2012). Assumer et risquer une posture de recherche collaborative. In Bourassa, B. and Boudjaoui, M. (dir.) *Recherches collaboratives en sciences humaines et sociales. Enjeux, modalités et limites*. Québec : Presses de l'Université du Québec.

Boyera, S., Koulla-Shirod, S., Abe, C., Spirea, B., Moattia, J-P. (2011). Implementing operational research to scale-up access to antiretroviral therapy for HIV infection: lessons learned from the Cameroonian experience. *Current Opinion in HIV and AIDS*, 6, 239-244.

Casabiaca, F. et Albeladejo, C. (1997). Des multiples légitimités de la recherche action. *Etud. Rech. Syst. Agraires Dev*, 30, 11-25.

Champagne, M. (2007). La pratique de la recherche-action. Entre utopie et nécessité. In Dorvill, H. (dir.) *Problèmes sociaux Tome III. Théorie et méthodologie de la recherche*, 463-490. Montréal : Presses de l'Université du Québec.

Chevalier, J., Buckless, D. J et Bourassa, M. (2013). *Guide de la recherche action, la planification et l'évaluation participative, Module 2*. SAS2. Ottawa, Canada.

D'Alonzo, K. T. (2010). Getting started in CBPR: lessons in bulding community partenerships for new researchers. *Nursing Inquiry*, 17(4), 282-288.

Demange, E., Henry E. et Préau M. (2012). *De la recherche en collaboration à la recherche communautaire. Un guide méthodologique*. Paris. ANRS/Coalition Plus. Coll. Sciences sociales et sida. 2012.

Desclaux, A., Kouanda, S., and Obermeyer, C.M. (2010). Stakeholders participation in operational research on HIV: insights from Burkina Faso. *AIDS*, 24(1), S79-85.

Dickens, L and Watkins, K. (1999). Action Research: Rethinking Lewin. *Management Learning*, 30, 127.

Ministère de l'économie et des finances. République du Burkina Faso (2011). *Stratégie de croissance accélerée et de développement durable* 2011 – 2015. http://www.unpei.org/sites/default/files/PDF/Burkina_Faso_PRSP_2011.pdf

Faure, G., Hocdé, H., et Chia, E. (2011). Action research methodology to reconcile product standardization and diversity of agricultural practices: A case of farmers' organizations in Costa Rica. *Action Research*, 9, 242-260.

Faure, G., Hocdé, H. et Meneses, D. (2007). Les organisations paysannes du Costa Rica construisent leur vision de l'agriculture familiale : une démarche de recherche-action marquée par une rupture. *Cahiers Agricultures*, 16(3), 205-2011.

Hercot, D., Meessen, B., Ridde, V. et Gilson, L. (2011). Removing user fees for health services in low- income countries: a multi-country review framework for assessing the process of policy change. *Health Policy and Planning*, 26, ii5–ii15.

Le May, A. et Lathlean, F. (2001). Action research: A design with potential. *Nursing Times Research*, 6, 502.

Lui, M. (1997). La validation des connaissances au cours de la recherche action. Etud. *Rech. Syst. Agraires Dev*, 30, 183-196.

McVicar, A., Munn-Giddings, C. et Abu-Helil, C. (2012). Exploring the development of action research in nursing and social care in the UK: A comparative bibliometric review of action research designs in social work (2000-2010). *Action Research*, 10, 79.

Mikolasek, O., Barlet, B., Chia, E., Pouomogne, V. et Tabi, M.T.E. (2009).

Développement de la petite pisciculture marchande au Cameroun: la recherche-action en partenariat. *Cahiers Agricultures*, 18(2-3), 270- 275.

Mshelia, C., Huss, R., Mirzoev, T., Elsey, H., Baine, S.O., Aikins, M., *et al.* (2013). Can action research strengthen district health management and improve health workforce performance? A research protocol. *BMJ Open* 3.

OMS. (2008). *Guide to operational research in programs support by the Global Fund*. Organisation mondiale de la santé. Fonds mondial pour la lutte contre Sida, la tuberculose et le paludisme.

Patton, M.Q. (2001). Evaluation, knowledge management, best practices, and high quality lessons learned. *American Journal of Evaluation*, 22, 329.

Peters, D. H., Tran, N. T. et Adam, T. (2013). *Implementation research in health: a practical guide*. Geneva: World Health Organization. Alliance for Health Policy and Systems Research
http://who.int/alliance-hpsr/alliancehpsr_irpguide.pdf

Peters, D.H., Taghreed, A., Alonge, O., Agyepong, I.A. et Tran, N. (2013). Implementation research: what it is and how to do it. *BMJ* 347.

Reason, P., et Bradbury, H. (2011). Introduction: Inquiry and participation in search of a world worthy of human aspiration. In P. Reason & H. Bradbury (Eds.), *Handbook of action research: Participative inquiry and practice* (pp. 1–14). London: SAGE.

Roy, M. et Prévost, P. (2013). La recherche-action : origines, caractéristiques et implications de son utilisation dans les sciences de la gestion. *Recherches qualitatives*, 32(2), 129-151.

Swann, C. (2002). Action research and the practice of design. *Design Issues*, 18(2).

Tousignant, P., Remondin, M., Laurendeau, M-C., et Bergeron, P. (2012). Une approche participative pour une recherche ancrée dans la pratique en première ligne. *Réseau de recherche en santé des populations du Québec. Carnets-synthèse* 11.
http://santepop.qc.ca/fr/activites/carnets.html

Vespiren, M. et Chia, E. (2012). Rôle d'une recherche-action sur la diffusion des savoirs et la modification du contexte social. Bourassa, B. and Boudjaoui, M. (dirs) *Recherches collaboratives en sciences humaines et*

sociales. Enjeux, modalités et limites. Montréal : Presses de l'Université du Québec.

Waterman, H., Marshall, M., Noble, J., Davies, H., Walshe, K., Sheaff, R. et Elwyn, G. (2007). The Role of Action Research in the Investigation and Diffusion of Innovations in Health Care: The PRIDE Project. *Qual Health Res*, 17, 373.

Zachariah, R., Harries, A.D., Ishikawa, N., Rieder, H.L., Bissell, K., Laserson, K., *et al.* (2009). Operational research in low-income countries: what, why, and how? *Lancet Infect Dis*, 9(11), 711- 717.

Zachariaha, R., Ford, N. Draguez, B., Yunc, O. et Reida, T. (2010). Conducting operational research within a non governmental organization: the example of Médecins Sans Frontières. *International Health* 2(1-8).

Résumé / Abstract / Resumen

Depuis 2011, une recherche-action est menée dans le cadre d'un projet de subvention de l'accès aux soins des enfants de moins de cinq ans, des femmes enceintes et allaitantes au district sanitaire de Sebba au Burkina Faso. Elle vise à expérimenter une innovation, le remboursement au forfait des prestations de soins de santé délivrés gratuitement aux patients afin de résoudre les difficultés inhérentes au projet de subvention de l'accès aux soins rencontrées par les agents de santé et l'ONG subventionnaire. Parmi les difficultés figurent entres autres la charge de travail pour le personnel de santé, les coûts de production des outils de collecte des données. Les résultats préliminaires de cette recherche-action étant probants, il est apparu nécessaire d'analyser la démarche utilisée pour la mettre en œuvre. Pour y parvenir, une analyse externe combinée à une analyse réflexive des parties prenantes a permis de tirer les leçons idoines qui ont permis de guider d'autres recherches-actions. La réussite d'une telle démarche d'analyse est passée par i) l'élaboration d'une grille d'analyse comportant des éléments factuels pour mener à bien l'analyse externe, ii) une bonne préparation de l'analyse réflexive et iii) une implication des parties prenantes au processus d'analyse dès le début.

***

Since 2011, action research has been conducted as part of a project to subsidize access to care for children under five, pregnant and lactating

women in Sebba health district in Burkina Faso. It aims to test an innovation, the reimbursement of health care benefits delivered free of charge to patients in order to solve the difficulties inherent to the project of grant of access to care met by the health workers and the subsidizing NGO. As a difficulty he had in other workload for the health staff, the costs of production of data collection tools. The preliminary results of this action research being conclusive, it appeared necessary to analyze the approach used to implement it. To achieve this, an external analysis combined with a reflective analysis of the stakeholders made it possible to draw the appropriate lessons that made it possible to guide other action research. The success of such an analysis has gone through i) the development of an analysis grid with factual elements to carry out the external analysis, ii) a good preparation of the reflexive analysis and iii) Stakeholder involvement in the analysis process from the beginning.

***

Desde 2011, una investigación-acción se lleva a cabo como parte de un proyecto para subvencionar el acceso a la atención de los niños menores de cinco años y de las mujeres embarazadas y lactantes en el distrito de salud de Sebba, en Burkina Faso. Su objetivo es experimentar una innovación, el reembolso de los servicios de asistencia sanitaria gratuitos prestados a los pacientes, con el fin de resolver las dificultades inherentes al proyecto de subvención del acceso a la asistencia encontradas por los trabajadores de la salud y de la ONG financiadora. Los desafíos incluyen la carga de trabajo del personal de salud, los costos de producir herramientas de recolección de datos, etc. Como los resultados preliminares de esta investigación-acción fueron concluyentes, parecía necesario analizar el enfoque utilizado para llevarla a cabo. Para lograrlo, un análisis externo combinado con un análisis reflexivo de las partes interesadas proporcionó las lecciones apropiadas para guiar futuras investigaciones-acciones. El éxito de este enfoque de análisis se ha logrado a través de (i) el desarrollo de una tabla de análisis con elementos factuales para llevar a cabo el análisis externo, (ii) una buena preparación del análisis reflexivo y (iii) la participación de las partes interesadas en el proceso de análisis desde el principio.

***

Aka Bony Roger Sylvestre est médecin de formation et titulaire d'un Master en santé publique option Santé mondiale en 2014 à l'Université de

Montréal au Canada avec orientation évaluation. Comme stagiaire au Centre de Recherche du Centre Hospitalier de l'Université de Montréal, il a conçu un outil et analysé le processus d'implantation d'une recherche-action en santé au Burkina Faso. Les résultats de ce travail ont fait l'objet de plusieurs publications notamment dans la revue de la Société Québécoise d'Évaluation de Programme au Canada et a remporté un prix d'excellente au concours « Évaluation d'Or 2015 » en Côte d'Ivoire. Depuis janvier 2018, il est chef d'un projet de recherche sur la tuberculose chez les enfants au Programme PACC-CI en Côte d'Ivoire. De 2015 à 2017, il était chargé de suivi-évaluation au Programme élargi de vaccination au sein du ministère de la Santé en Côte d'Ivoire.

Valéry Ridde est directeur de recherche au CEPED[1], une Unité Mixte de Recherche associant l'Université Paris Descartes et l'Institut de Recherche pour le Développement (IRD). Il est chercheur régulier de l'Institut de recherche en santé publique de l'Université de Montréal (IRSPUM) et chercheur associé au Centre interdisciplinaire de recherche en développement international et société (CIRDIS) de Montréal. Il est rédacteur adjoint des revues *BMJ Global Health* et *Global Health Promotion*, membre du comité de rédaction de la *Revue Canadienne d'Évaluation de Programme* et de *Qualitatif! Revue africaine de sciences sociales*. Ses travaux de recherche portent sur la couverture universelle en santé, le financement des services de santé, l'évaluation de programme, les politiques publiques de santé et la promotion de la santé.

Ludovic Queuille est Consultant en Gouvernance & Financement de la Santé pour l'Organisation panaméricaine de la santé (OPS)/Organisation mondiale de la santé (OMS) en Haïti depuis septembre 2014. C'est en tant que Courtier en connaissances (2008-2014) pour le compte du Centre de recherche de l'Université de Montréal et basé à Ouagadougou qu'il était, entre autres travaux, responsable de cette recherche-action sur le paiement par montants forfaitaires dans le cadre d'un projet de subvention du paiement de soins des enfants, femmes enceintes et allaitantes dans la région du Sahel au Burkina Faso (http://www.equitesante.org/helpburkina/).

1. http://www.ceped.org

Citation

Aka Bony Roger Sylvestre, Valéry Ridde et Ludovic Queuille. (2019). L'analyse d'une recherche-action. Combinaison d'approches dans le domaine de la santé au Burkina Faso. In *Évaluation des interventions de santé mondiale. Méthodes avancées*. Sous la direction de Valéry Ridde et Christian Dagenais, pp. 125-153. Québec : Éditions science et bien commun et Marseille : IRD Éditions.

PARTIE III

LES MÉTHODES MIXTES

6. Les revues systématiques mixtes

Un exemple à propos du financement basé sur les résultats

QUAN NHA HONG, ANNE-MARIE TURCOTTE-TREMBLAY ET PIERRE PLUYE

Les revues systématiques mixtes

Définition de la méthode

Les revues systématiques mixtes sont des revues de littérature utilisant une approche systématique pour combiner des études quantitatives, qualitatives et mixtes.

Forces de la méthode

- Utilise une démarche explicite, transparente et reproductible
- Combine des études quantitatives, qualitatives et mixtes
- Permet d'expliquer des phénomènes complexes et multidimensionnels
- Fournit un portrait plus complet de la littérature

Défis de la méthode

- Requiert du temps et des ressources pour minimiser les biais de détection, sélection, divulgation et interprétation
- Nécessite une intégration des données quantitatives et qualitatives

Les dernières années ont été marquées par un développement considérable des méthodes de synthèse des études et par une multiplication des types de revue, incluant les revues systématiques mixtes. Ce chapitre décrit brièvement les étapes de cette méthode en mettant l'accent sur la synthèse des données, soit une des étapes de la revue qui consiste à analyser les données pour en faire un tout cohérent (Mays, Pope et Popay 2005).

Les revues mixtes de littérature utilisent une approche systématique et incluent des études utilisant des méthodes quantitatives, qualitatives et/ou mixtes (Pluye et Hong 2014). D'autres termes sont utilisés pour dénommer ce type de revue : revue intégrative (*integrative review*) (Whittemore et Knafl 2005) ou revue utilisant des méthodes mixtes (*mixed methods review*) (Harden 2010). Il n'existe donc pas encore de consensus sur la terminologie. Dans ce chapitre, nous utiliserons le terme « revue systématique mixte ».

La revue systématique mixte se base sur les principes des méthodes

mixtes, c'est-à-dire l'intégration des phases, résultats et données de type qualitatif et quantitatif afin de tirer profit de leur complémentarité (Creswell *et al.* 2011).

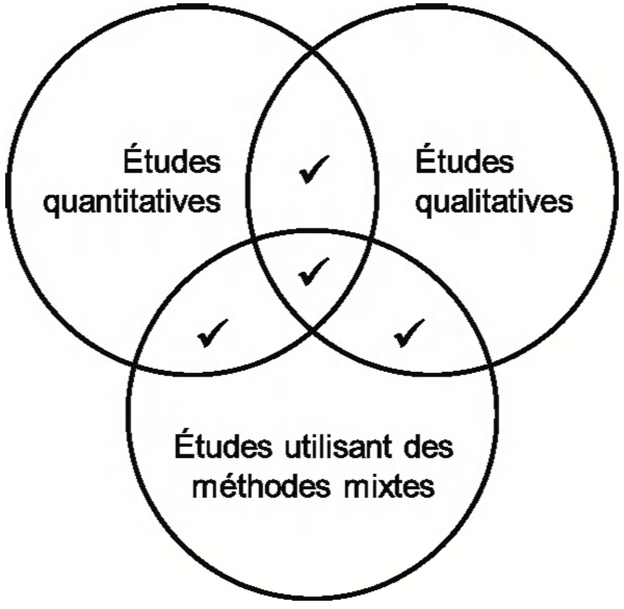


Figure 1. *Types d'études inclus dans une revue systématique mixte*

De façon générale, il est possible d'identifier trois principaux types d'études primaires, soit les études quantitatives, les études qualitatives et les études mixtes. Comme le montrent les crochets dans les zones d'intersection de la figure 1, une revue combinant deux ou trois types d'études primaires peut être considérée comme mixte tant et aussi longtemps que des données qualitatives et quantitatives sont utilisées (Pluye et Hong 2014). Aussi, comme les études mixtes sont caractérisées par la combinaison de méthodes qualitatives et quantitatives (Creswell et al. 2011), une revue de ces études pourrait également être considérée comme mixte lorsque les données qualitatives et quantitatives sont synthétisées. Autrement dit, la revue de différentes études au sein d'une même catégorie n'est pas mixte. Ainsi, combiner les résultats de différents types d'étude quantitative (par exemple, des essais randomisés et non randomisés) constitue une revue quantitative, alors que combiner ceux de différents

types d'étude qualitative (par exemple, étude de cas et ethnographie) génère une revue qualitative.

En combinant des études qualitatives, quantitatives et mixtes, la revue systématique mixte permet d'aborder des questions complexes et d'approfondir la compréhension des phénomènes observés. Elle permet d'obtenir (i) une meilleure compréhension des résultats des études quantitatives, (ii) une généralisation des résultats des études qualitatives, fondée sur des inférences statistiques ou (iii) une corroboration de connaissances obtenues par des méthodes quantitatives et qualitatives (Pluye *et al.* 2009). Aussi, la revue systématique mixte peut fournir un portrait plus large et complet de la littérature sur un sujet donné, notamment en combinant des questions complémentaires telles que : Quelle est l'efficacité d'une intervention? Pourquoi cette intervention est-elle efficace ou non? Comment l'intervention fonctionne-t-elle? Quels sont les barrières et les facilitateurs pour implanter cette intervention? Les réponses à ces questions complémentaires peuvent avoir des implications pratiques, telles que favoriser la prise de décisions chez les décideurs-ses politiques, les gestionnaires et les clinicien-ne-s.

Plusieurs taxonomies des revues de littérature ont été proposées. Elles ont identifié une diversité de types de revues variant selon plusieurs critères tels que les objectifs poursuivis, les méthodes (étapes) utilisées et l'audience visée (Cooper 1988; Grant et Booth 2009; Paré *et al.* 2015). La revue systématique mixte est considérée comme le type le plus rigoureux, car il permet de minimiser plusieurs biais liés à la sélection et à l'analyse des études.

Toutes les étapes de la revue systématique doivent être transparentes et explicites faisant en sorte qu'elle puisse être reproduite (Pluye *et al.* 2016). La revue systématique vise à répondre à une question clairement définie en suivant une méthode systématique et explicite dans le repérage des études pertinentes, dans leur évaluation et dans leur synthèse (Chalmers, Hedges et Cooper 2002). Dans une revue systématique, les sources d'information sont multiples. La recherche documentaire est conduite par des bibliothécaires spécialisés et est exhaustive (incluant la littérature grise). La sélection, l'évaluation de la qualité des études incluses et l'extraction des données sont effectuées par au moins deux spécialistes de manière indépendante.

Application de la méthode

Cette section vise à présenter les étapes à suivre pour mener une revue systématique mixte. Pour illustrer ces étapes, nous prenons l'exemple d'une revue systématique mixte sur les conséquences non intentionnelles d'une intervention reposant sur le financement basé sur les résultats (FBR). L'encadré 1 décrit ce qu'est le FBR et explique la pertinence de la revue systématique mixte dans ce domaine.

Encadré 1. Financement Basé sur les Résultats (FBR)

Le financement basé sur les résultats (FBR) est une intervention visant à offrir des incitatifs financiers aux travailleurs et travailleuses de la santé pour améliorer la qualité et la quantité des soins. Cette intervention a été développée dans les pays à revenu élevé et s'est rapidement diffusée dans les pays à faible et à moyen revenu (PFMR) avec le soutien d'organisations internationales investies dans la réduction de la mortalité maternelle et infantile. Plusieurs appellations et modèles d'intervention du FBR existent. Selon le modèle de la Banque Mondiale (Fritsche, Soeters, et Meessen 2014), des ententes de rendement sont établies pour définir les indicateurs et les objectifs de performance mesurables et vérifiables. Les établissements de santé et/ou les travailleurs et travailleuses de la santé sont payés pour chaque service de santé prédéterminé. Des bonus sont également offerts selon le niveau de qualité des soins. Des examinatrices et examinateurs indépendants procèdent à des vérifications, notamment des registres et des lieux, pour promouvoir la transparence et la reddition de comptes.

Un corpus grandissant d'études qualitatives et quantitatives suggère que le FBR peut engendrer, au-delà des effets positifs attendus, des conséquences non intentionnelles, c'est-à-dire des changements désirables ou indésirables pour lesquels il y a un manque d'action ou de causalité intentionnelle et qui se produisent dans un système social à la suite d'une innovation. Pour synthétiser les résultats des études existantes et complémentaires sur le FBR, nous effectuerons une revue systématique mixte.

Nous avons sélectionné ce type de revue en raison de sa capacité à prendre en compte l'hétérogénéité des études et la complexité des interventions. La littérature sur le FBR se caractérise par une multiplicité de contextes d'implantation, de méthodologies, de méthodes, de modèles d'intervention et de résultats de recherche. La synthèse de ces études complémentaires est nécessaire pour acquérir une compréhension globale de l'état des connaissances sur cet enjeu complexe. Les résultats de cette revue systématique mixte permettront, d'une part, de créer une grille d'analyse préliminaire dans le cadre d'une recherche sur les conséquences

non intentionnelles du FBR au Burkina Faso et, d'autre part, d'aider la prise de décision sur l'implantation du FBR.

Comme présenté dans le tableau 1, les revues systématiques suivent principalement huit étapes. Les étapes de la revue systématique sont similaires à celles des études primaires avec comme principale différence que l'unité d'analyse est une étude (et non pas un-e participant-e ou une organisation).

Tableau 1. Étapes de la revue systématique et comparaison avec les étapes d'une étude primaire

Revue systématique	**Étude primaire**
Formulation d'une question de recherche	Formulation d'une question de recherche
Définition des critères d'éligibilité	Définition des critères d'éligibilité
Identification des sources de documents	Identification des sources de données
Application d'une recherche documentaire	Recrutement des participant-e-s
Sélection des études pertinentes	Sélection des participant-e-s
Évaluation de la qualité des études retenues	S.O.
Extraction des données des études retenues	Collecte des données
Synthèse des données extraites	Analyse des données

Dans ce qui suit, les étapes de la revue systématique sont brièvement décrites et illustrées à partir du cas du FBR. Une liste de références méthodologiques sur la revue systématique est fournie à la fin de ce chapitre.

Étape 1 : Formuler une question de recherche

Dans les revues systématiques mixtes, il peut y avoir une ou plusieurs questions de recherche documentaires. Ces questions doivent être claires et précises afin de bien cibler le sujet. Pour notre exemple de cas sur le FBR, nous avons formulé la question d'intérêt suivante : « Quelles sont les conséquences non intentionnelles du FBR dans les pays à faible et à moyen revenu (PFMR)? ». Cette question permet d'intégrer des résultats de nature qualitative et quantitative. Nous aurions également pu ajouter une question complémentaire comme « Quelle est l'importance respective de ces conséquences non intentionnelles? » ou « Quelles stratégies sont implantées pour faire face à ces conséquences non intentionnelles? ». Notre choix concernant la formulation de la question a été guidé par les ressources disponibles ainsi que l'utilité et la pertinence. La formulation de la question, ou des questions, est importante, car cela influence toutes les étapes subséquentes de la revue.

Étape 2 : Définir les critères d'éligibilité

Les critères d'éligibilité sont associés aux concepts-clés dans la question de recherche et sont constitués des critères d'inclusion et d'exclusion. Le tableau 2 présente les critères établis à partir de notre question de recherche sur le FBR. Il est important de clairement définir et justifier chaque critère. Par exemple, nous avons décidé d'exclure les études publiées avant l'année 2000 étant donné que le FBR s'est grandement répandu dans les pays à faible et à moyen revenu depuis les années 2000. Aussi, nous limitons la sélection aux études scientifiques, c'est-à-dire composées d'une question/objectif de recherche, d'une collecte/analyse de données (issues d'expérimentation, d'observation ou de simulation) et de résultats. La combinaison d'études qualitatives et quantitatives augmente le nombre d'articles à synthétiser. En conséquence, il est nécessaire de restreindre le plus possible les critères de sélection, c'est-à-dire de bien cibler la ou les questions.

Tableau 2. Critères d'inclusion et d'exclusion

	Critère d'inclusion	Critère d'exclusion
Problème	La faible qualité et quantité des services de santé	Problèmes non reliés aux services de santé
Intervention	Financement basé sur les résultats (incitatifs à l'offre)	Aucun financement basé sur les résultats ou interventions offrant des incitatifs aux patient-e-s ou aux compagnies d'assurance
Résultats	Conséquences non intentionnelles	Conséquences intentionnelles seulement
Lieu	Pays à faible et à moyen revenu	Pays à revenu élevé
Type d'étude	Études scientifiques utilisant des méthodes qualitatives, quantitatives ou mixtes.	Études non scientifiques, protocoles de recherche, éditoriaux, lignes directrices (c.-à-d., texte sans collecte ou analyse de données suivant une méthode de recherche)
Langues	Anglais, français et espagnol	Articles écrits dans une langue autre que l'anglais, le français et l'espagnol
Années de publication	2000 – 2016	Avant 2000

Étape 3 : Sélectionner les sources d'information

Pour s'assurer de repérer l'ensemble des études pertinentes à la question de recherche, diverses sources de données doivent être utilisées. Pour cette étape et la suivante, la collaboration avec un(e) ou des bibliothécaires maîtrisant les techniques de recherche dans les bases de données et d'autres types de sources est fortement recommandée. Pour la revue systématique mixte sur le FBR, nous avons sélectionné cinq bases de données bibliographiques complémentaires dans les domaines de la santé (MEDLINE, CINAHL, EMBASE), des sciences humaines, sociales et économiques (FRANCIS, Business Source Premier). D'autres sources d'information peuvent également être utilisées pour s'assurer d'avoir une recherche exhaustive. Pour notre revue sur le FBR, nous avons utilisé quatre autres sources d'information :

(a) consulter la liste des articles retenus dans des revues systématiques

sur le FBR (Gorter et Meessen 2013; Witter *et al.* 2012) et la liste de références des articles retenus

(b) chercher la littérature grise en consultant des sites Internet dans le domaine (par exemple, www.rbfhealth.org, www.healthfinancingafrica.org, www.theglobalfund.org) et des thèses dans la base Proquest Dissertations & Theses Global

(c) envoyer la liste des articles retenus à des chercheuses et chercheurs ayant une expertise sur le FBR afin qu'ils et elles identifient des articles pertinents manquants

(d) effectuer une recherche par citations (c'est-à-dire chercher les articles qui ont cité chaque article retenu) dans une base de données bibliographique spécialisée comme Scopus ou Google Scholar.

Étape 4 : Élaborer une stratégie de recherche documentaire

Pour élaborer une stratégie de recherche documentaire, il est nécessaire d'identifier les termes reliés aux concepts-clés dans la question de recherche. Pour la revue systématique mixte sur le FBR, nous avons élaboré une stratégie de recherche documentaire autour de trois principaux concepts, soit les conséquences non intentionnelles, le financement basé sur les résultats et les pays à faible et à moyen revenu (voir un exemple au tableau 3). Cette stratégie a été adaptée pour chaque base de données retenue. Une fois la recherche documentaire terminée, toutes les références doivent être exportées dans un logiciel de gestion documentaire tel que Zotero, EndNote, RefWork ou Mendeley. Étant donné que la recherche est effectuée dans plusieurs bases de données qui se chevauchent, une même référence peut apparaître plusieurs fois dans le logiciel (doublons). Ces doublons sont identifiés et enlevés en suivant une méthode de déduplication, comme la méthode de Bramer *et al.* (2016) ou de Rathbone *et al.* (2015). Ces méthodes permettent d'identifier les doublons en choisissant des champs spécifiques (par exemple, auteur, titre, année, journal, page, volume et numéro) selon un ordre précis.

Tableau 3. Exemple d'une stratégie de recherche documentaire dans PUBMED

Search	Query
#1	Search unintend* OR intend* OR unpredict* OR predict* OR unanticipat* OR anticipat* OR spillover* OR indirect* OR direct* OR unaware* OR aware* OR unexpect* OR expect* OR desirable* OR Undesirable* OR "cherry picking" OR "cream skimming" OR gaming* OR inequit* OR equit* [ALL]
#2	Search "pay for performance" OR "pay for value" OR "pay for quality" OR "P4P" OR "PBF" OR "value-based purchasing" OR ("monetary incentive" AND quality) OR ("monetary incentives" AND quality) OR ("financial incentive" AND quality) OR ("financial incentives" AND quality) OR (bonus AND quality) OR (bonuses AND quality) OR (reward AND quality) OR (rewards AND quality) OR "performance-based payment" OR "performance-based pay" OR "performance-based reimbursement" OR "performance-based financing" OR "performance-based contracting" OR "results-based financing" OR "output-based payment" OR "incentive reimbursement" OR "incentive reimbursements" OR "quality-based purchasing" OR "quality incentive" OR "quality incentives" [TIAB]
#3	Search « Developing Countries » OR "Africa" OR "Africa, Northern" OR "Africa South of the Sahara" OR "Africa, Central" OR "Africa, Eastern" OR "Africa, Southern" OR "Africa, Western" OR "Asia" OR "Asia, Central" OR "Asia, Southeastern" OR "Asia, Western" OR "Caribbean Region" OR "West Indies" OR "South America" OR "Latin America" OR Central America [MH]
#4	Search #1 AND #2 AND #3
#5	Limit #4 to ([english OR french OR spanish [LA]] AND [2000:2016 [DPJ]])

Des chercheuses et chercheurs s'intéressent à développer et tester des filtres de recherche pour identifier des documents pertinents sur différents concepts. Par exemple, un filtre a été créé dans Medline pour identifier les études quantitatives, qualitatives et mixtes (El Sherif *et al.* 2016) tandis qu'un autre filtre permet d'identifier les documents portant sur les pays à faible et à moyen revenu (Cochrane Effective Practice and Organisation of Care 2012). L'utilisation de filtres de recherche permet de faciliter l'élaboration des stratégies de recherche documentaire et de mieux cibler les documents potentiellement pertinents.

Étape 5 : Sélectionner les études

Cette étape consiste à filtrer les résultats de la recherche documentaire pour sélectionner les études à inclure dans la revue systématique mixte. La

sélection des études se déroule généralement en deux parties. En premier lieu, les titres et les résumés sont lus pour juger de l'éligibilité des documents à partir des critères préétablis à l'étape 2. En second lieu, les textes sont lus dans leur intégralité. Dans une revue systématique, la sélection de l'ensemble du matériel est effectuée par deux évaluateurs et évaluatrices de façon indépendante afin de réduire les biais de sélection. En cas de désaccord, les évaluateurs et évaluatrices se rencontrent pour établir un consensus et une troisième personne est impliquée pour trancher si un désaccord persiste après discussion. L'accord interjuge peut être estimé en utilisant le Kappa de Cohen. À l'issue de ce processus de sélection, un diagramme de flux précisant le nombre de documents retenus et exclus à chaque étape ainsi que les raisons d'exclusion est présenté (figure 2) (adapté des lignes directrices PRISMA de Moher *et al.* (2009), traduit dans Gedda [2015]). La création du diagramme de flux requiert une grande organisation dès le départ. Tout le processus doit être documenté (principe de transparence).

Pour réduire le nombre de titres et résumés à filtrer, un classifieur automatisé de textes sur les études scientifiques (CATES) a été développé suite à une étude comparant la performance de divers classifieurs automatiques (Langlois *et al.* 2018). Ce classifieur permet de catégoriser automatiquement les publications indexées en (1) études scientifiques (qualitatives, quantitatives et mixtes) et (2) autres types de travaux (par exemple, commentaire, éditorial, revue de la littérature, article méthodologique, description de programme et ligne directrice). Le CATES est disponible à cette adresse internet et peut être utilisé gratuitement : https://babel.iro.umontreal.ca/.

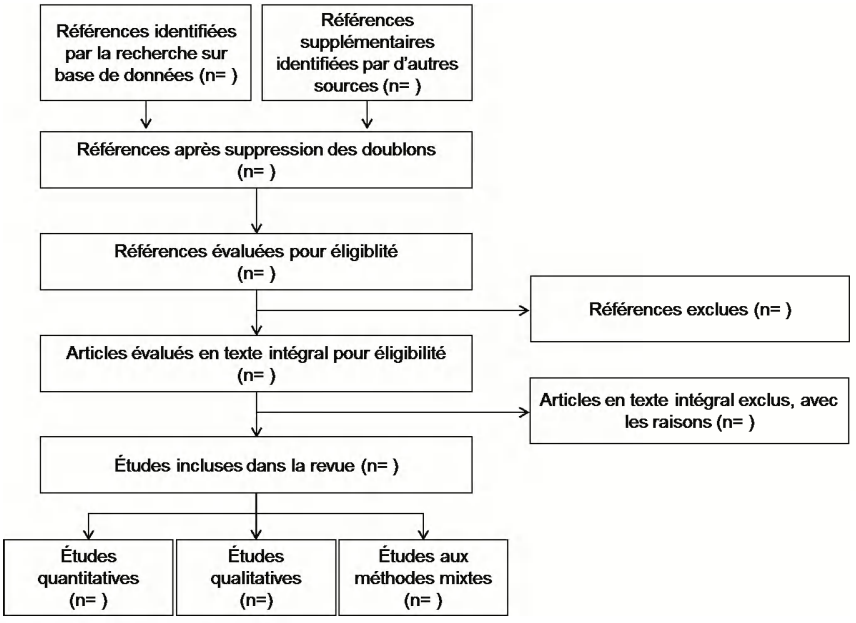


Figure 2. Diagramme de flux

Étape 6 : Évaluer la qualité des études

Dans une revue systématique, il est important d'évaluer la qualité des études incluses en prenant en considération les biais potentiels pour pouvoir tirer des conclusions valides sur l'état des connaissances. L'évaluation de la qualité des études est généralement effectuée en utilisant des outils d'évaluation (*critical appraisal tools*). Il existe une variété d'outils pour diverses méthodes de recherche (Bai *et al.* 2012; Crowe et Sheppard 2011; Heyvaert *et al.* 2013). Pour la revue systématique mixte sur le FBR, la qualité des études est évaluée à l'aide d'un outil développé spécifiquement pour ce type de revue, le *Mixed Methods Appraisal Tool* (MMAT) (Pluye *et al.* 2011; Pluye *et al.* 2009; Souto *et al.* 2015; Pace *et al.* 2012). Le MMAT permet d'évaluer de façon concomitante cinq types d'études : études qualitatives, essais contrôlés randomisés, études quantitatives analytiques, études quantitatives descriptives et études mixtes (Pluye *et al.* 2009). La dernière version de cet outil comporte 25 items sur la qualité méthodologique des études qui sont cotés sur une échelle nominale (oui/non/ne sais pas) (Hong

et al. 2018). L'évaluation de la qualité sert à décrire les articles retenus et à interpréter les données dans la synthèse. Il est possible de télécharger cet outil à partir de ce site internet : http://mixedmethodsappraisaltoolpublic.pbworks.com/.

Dans la revue FBR, les résultats de cette évaluation ne sont pas utilisés pour éliminer les études de faible qualité méthodologique, car elles peuvent quand même contribuer aux connaissances. Leurs limites sont prises en compte lors d'analyses de sensibilité. Les analyses de sensibilité permettent d'explorer comment les résultats obtenus dans la synthèse sont modifiés selon la qualité des études (Carroll et Booth 2015). Ainsi, ces analyses permettent de vérifier si les résultats de la synthèse sont similaires lorsque des études de faible qualité sont retirées. Par exemple, une similarité permet d'appuyer les résultats avec un plus grand nombre d'études et de rendre les résultats plus exacts pour les décideuses et décideurs.

Étape 7 : Extraire des données

Cette étape vise à collecter les données (résultats des études incluses ou données primaires lorsque ces dernières sont accessibles) à l'aide d'un formulaire d'extraction précisant toutes les données nécessaires. Les données extraites servent à décrire les études incluses (tableau des caractéristiques), décrire les résultats de ces études (tableau des résultats) et en effectuer la synthèse.

Un cadre conceptuel ou un modèle théorique est nécessaire pour guider la sélection des données à extraire, sauf si l'objectif de la revue consiste à construire un tel cadre ou modèle. Pour notre revue systématique mixte sur le FBR, nous avons développé un cadre conceptuel basé sur la théorie de diffusion des innovations de Rogers (2003). Cette théorie suggère que l'interaction entre plusieurs variables comme les caractéristiques des acteurs et actrices, la nature du système social ainsi que la nature et l'utilisation d'une nouvelle intervention peuvent engendrer différentes conséquences à travers le temps. Ainsi, nous extrayons les données suivantes des articles retenus : auteur, année, pays, méthodologie et méthodes, échantillon (population et nombre), nom de l'intervention FBR, conséquences non intentionnelles, nature et utilisation du FBR, caractéristiques des acteurs et actrices et nature du système social. Si des données sont

manquantes dans les articles analysés, leur premier auteur ou leur première autrice est contactée. Deux personnes font l'extraction de façon indépendante afin de s'assurer d'une compréhension uniforme des données et de diminuer les sources de biais et d'erreurs.

Étape 8 : Synthétiser les données extraites

La synthèse vise à analyser et combiner les données extraites et évaluées en un tout cohérent. Dans le cadre de la revue systématique mixte, la synthèse est plus complexe en raison de l'inclusion de données qualitatives et quantitatives. Différentes façons de synthétiser les données qualitatives et quantitatives peuvent être adoptées. L'encadré 2 présente les définitions des construits nécessaires pour bien comprendre les différentes façons de synthétiser les données.

Encadré 2 : Définitions de construits utilisés dans les devis de synthèse

Données : Dans une revue, les données sont les résultats des études incluses (ou les données primaires lorsque ces dernières sont accessibles). Les données peuvent être qualitatives (QL) (par exemple : des thèmes identifiés dans des entrevues) ou quantitatives (QT) (par exemple : des pourcentages identifiés dans un sondage).

Synthèse : Dans une revue, la synthèse vise à analyser les données. Il existe une variété de méthodes de synthèse qui peuvent être définies en deux principales catégories. La première regroupe les méthodes de synthèse quantitative (QT) permettant d'analyser des données sur une variable spécifique en utilisant des méthodes statistiques de base ou plus avancées (par exemple, méta-analyse, synthèse bayésienne et méta-régression). La seconde regroupe les méthodes de synthèse qualitative (QL) qui consiste à analyser des données pour générer comme résultats des catégories, thèmes, construits ou théories (par exemple, méta-ethnographie, synthèse thématique et synthèse réaliste). À l'heure actuelle, il est possible d'identifier plus d'une vingtaine de méthodes de synthèse qualitatives et quantitatives (Barnett-Page et Thomas 2009; Dixon-Woods *et al.* 2005; Popay 2006; Sutton et Higgins 2008;, Tricco *et al.* 2016). Dans une revue systématique mixte, les méthodes de synthèse utilisées dans une même revue peuvent être uniquement QL, uniquement QT, ou une combinaison des deux (QL et QT).

Résultats : Dans une revue, les résultats sont ceux issus de la synthèse des données. En fonction des méthodes de synthèse utilisées, les résultats de la revue peuvent être QL (par exemple, des thèmes ou un cadre conceptuel), ou QT (par exemple, des statistiques descriptives ou une taille d'effet), ou les deux (QL et QT).

Discussion : La discussion est une section qui vise à interpréter les résultats obtenus et à les comparer avec les principales idées publiées dans des articles de recherche ou non (par exemple, éditoriaux ou commentaires ou livres influents, guides de pratique, énoncés de politique publique, rapports institutionnels, etc.). Également, dans cette section, il y a généralement une présentation des limites de la revue et des recommandations.

Séquence : La séquence est définie selon une perspective de développement entre des phases de synthèse. Dans une revue mixte, une séquence est présente lorsque les résultats d'une phase informent la synthèse d'une phase subséquente.

Intégration : L'intégration est définie comme la combinaison des composantes quantitatives et qualitatives au sein d'une même étude (Plano Clark et Ivankova 2015). Dans une revue mixte, l'intégration peut survenir au niveau des données, des résultats de synthèse, de l'interprétation ou entre les phases.

Devis de synthèse

Une analyse de 459 revues mixtes a été effectuée pour étudier comment les données qualitatives et quantitatives sont analysées et suggérer des devis de synthèse (Hong *et al.* 2017). Inspirée de la littérature sur les méthodes mixtes (Creswell *et al.* 2011), l'analyse a été effectuée sur les méthodes de synthèse utilisées, la séquence et l'intégration des données qualitatives et quantitatives. Deux principaux types de devis de synthèse ont été identifiés : convergent et séquentiel (figure 3). Ces devis permettent de comprendre

comment la synthèse peut être effectuée dans le cadre d'une revue systématique mixte. Pour chacun de ces devis, il est possible d'utiliser des méthodes de synthèses qui sont uniquement qualitatives, uniquement quantitatives ou bien de combiner des méthodes de synthèses quantitatives et qualitatives (approche de synthèse mixte). Dans ce qui suit, chacun de ces devis est présenté.

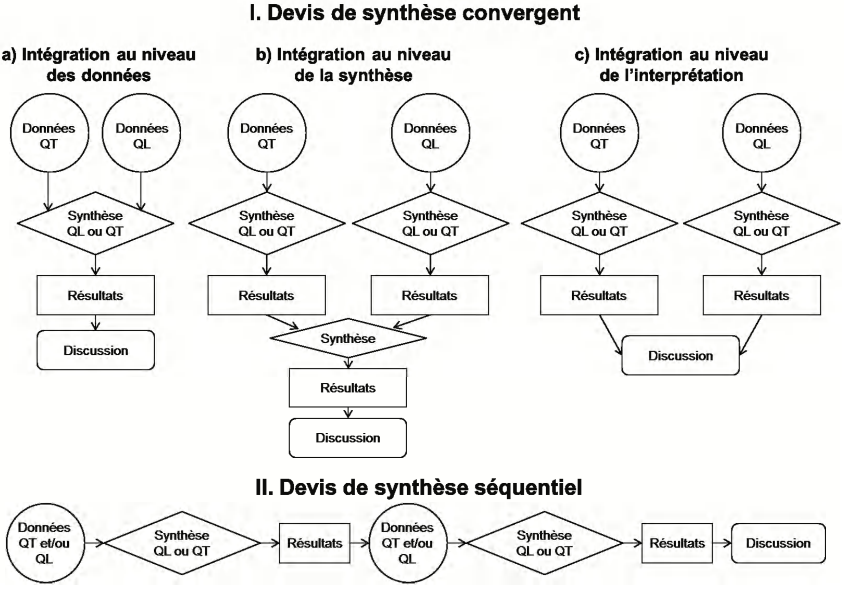


Figure 3. Devis de synthèse dans les revues systématiques mixtes

Devis de synthèse convergent

Dans un devis de synthèse convergent, les études qualitatives et quantitatives sont analysées de manière simultanée. Au sein de ces devis, il est possible d'identifier trois niveaux d'intégration, soit au niveau des données, des résultats de synthèse, ou de l'interprétation.

Devis de synthèse convergent avec intégration au niveau des données

Toutes les études sont analysées en utilisant une même méthode de

synthèse qui peut être uniquement qualitative ou quantitative (figure 3). Cela implique une transformation de données, c'est-à-dire que des données quantitatives sont transformées en catégories ou thèmes (synthèse qualitative), ou des données qualitatives sont transformées en variables et valeurs numériques (synthèse quantitative).

Par exemple, dans ce devis de synthèse, la question de recherche de notre exemple de cas serait « Quelles sont les conséquences non intentionnelles du FBR dans les pays à faible et à moyen revenu? ». Pour répondre à une telle question, une méthode de synthèse qualitative peut être utilisée, par exemple une synthèse thématique qui consiste à extraire les principaux thèmes de toutes les études retenues (Thomas et Harden 2008). Dans cet exemple, la transformation consiste à interpréter les données dans les études quantitatives (fréquences, tailles d'effets) de façon qualitative (en termes de thèmes).

Si un chercheur ou une chercheuse voulait fournir des statistiques sur les conséquences non intentionnelles du FBR, il lui faudrait, par exemple, effectuer une analyse de contenu quantitative qui consiste à attribuer des codes (quelques variables clés et valeurs numériques associées) aux résultats des études à partir d'une grille de codification validée et fidèle (Neuendorf 2002). Ainsi, les codes attribués aux études pourront être analysé de façon statistique. Il y a donc transformation car des valeurs numériques seront attribuées aux données issues des études qualitatives.

Devis de synthèse convergent avec intégration au niveau des résultats de synthèse

Les données qualitatives et quantitatives sont analysées et présentées séparément, mais intégrées en utilisant une autre méthode de synthèse (figure 3). L'intégration peut consister en une comparaison ou une juxtaposition des résultats des deux synthèses à l'aide de tableaux ou de matrices ou bien d'une analyse secondaire ou en sous-groupes.

Dans ce devis de synthèse, il y a habituellement une question de recherche générale et des sous-questions. Par exemple, pour répondre à la question « Quelles sont les conséquences non intentionnelles du FBR? », il peut y avoir deux sous-questions « Quelles sont les conséquences non

intentionnelles sur la satisfaction des soignant-e-s? » et « Quelles sont les conséquences non intentionnelles sur la quantité des soins de santé prodigués? ».

Pour répondre à ces questions, trois synthèses combinant des méthodes qualitatives et quantitatives pourraient être requises : (a) une synthèse qualitative des conséquences non intentionnelles sur la satisfaction des soignant-e-s, (b) une synthèse quantitative de type méta-analyse des essais randomisés sur la quantité des soins de santé et (c) une juxtaposition des résultats des deux synthèses pour examiner les liens entre la satisfaction des soignants et la quantité des soins offerts.

Devis de synthèse convergent avec intégration au niveau de l'interprétation

Les données qualitatives et quantitatives sont synthétisées et présentées séparément. L'intégration a lieu seulement lors de l'interprétation des résultats dans la discussion (figure 3).

Dans ce devis de synthèse, des questions de recherche distinctes et complémentaires sont généralement abordées. Par exemple, « Est-ce que le FBR est efficace? », « Quel est le coût-efficacité du FBR? », « Quelles sont les conséquences non intentionnelles du FBR? » et « Quels sont les enjeux éthiques reliés au FBR? ».

Pour répondre à ces questions de recherche complémentaires, des méthodes de synthèses qualitatives et quantitatives sont nécessaires. Par exemple, des synthèses quantitatives peuvent répondre aux deux premières questions sur l'efficacité et le coût-efficacité. Des synthèses qualitatives peuvent analyser les études qualitatives sur les enjeux éthiques et les conséquences non intentionnelles. La discussion résumera l'ensemble des résultats obtenus sur l'efficacité, le coût-efficacité, les conséquences inattendues et les enjeux éthiques. De plus, elle fournira des pistes de recommandations sur l'évaluation et l'implantation des FBR. Ce type de devis est observé, entre autres, dans les rapports d'évaluation des technologies de la santé où l'évaluation doit tenir compte de diverses dimensions (par exemple, efficacité, efficience, sécurité, impacts éthiques, psychologiques, organisationnels et économiques) (Busse *et al.* 2002).

Devis de synthèse séquentiel

Un devis de synthèse séquentiel implique au moins deux phases de synthèse dépendantes. Les résultats d'une première phase où les données qualitatives et/ou quantitatives sont synthétisées informent une seconde phase de synthèse (figure 3).

Dans un devis de synthèse séquentiel, il peut y avoir une question de recherche ou des questions dépendantes. Par exemple, les questions peuvent être « Quelles sont les conséquences non intentionnelles du FBR? » et « Quelle est l'importance relative de ces conséquences non intentionnelles? ». Pour répondre à ces questions, deux phases sont requises. Une première synthèse qualitative (phase 1) permettra d'établir une liste de conséquences non intentionnelles. À partir de cette liste (résultats de la phase 1), les études quantitatives sont analysées avec une méthode de synthèse quantitative (phase 2) pour déterminer l'importance relative des conséquences non intentionnelles. Ceci permettra de quantifier l'importance des conséquences non intentionnelle ainsi que d'identifier les lacunes dans les connaissances, c'est-à-dire les conséquences qui n'ont pas été étudiées dans les études quantitatives.

Les derniers exemples présentés utilisent une approche de synthèse mixte (c'est-à-dire l'utilisation des méthodes de synthèse qualitatives et quantitatives dans une même revue). Selon les questions de recherche, il est possible d'utiliser seulement des méthodes de synthèse qualitative ou bien seulement des méthodes de synthèse quantitative. Par exemple, des méthodes de synthèse qualitative peuvent être utilisées dans les deux phases d'un devis de synthèse séquentiel. Dans une première phase, une méthode de synthèse inductive est utilisée pour définir les principaux thèmes. Puis dans une deuxième phase, les études sont analysées de façon déductive à partir des thèmes identifiés.

Analyse réflexive

La force d'une revue systématique est sa démarche explicite, transparente et reproductible. Plusieurs lignes directrices pour rapporter une revue systématique ont été élaborées dont la plus connue est le PRISMA (Moher *et al.* 2009).

Plusieurs biais de recherche peuvent être minimisés en suivant cette démarche. Un premier biais potentiel est celui de la détection qui est relié à la recherche documentaire. Afin de s'assurer de repérer l'ensemble des études pertinentes pour la question de recherche (recherche exhaustive), il est important de diversifier les sources documentaires, d'utiliser plusieurs bases de données et de développer des stratégies de recherche documentaire rigoureuse avec des bibliothécaires spécialisé-e-s. Un deuxième biais est celui de la sélection. Il est important d'éviter d'avoir une sélection arbitraire des études en définissant des critères de sélection clairs et d'impliquer deux personnes dans la sélection. Le biais de divulgation des résultats est un troisième biais à considérer ; le plus connu est le biais de publication. En effet, une étude avec des résultats positifs a plus de chance d'être publiée et plus rapidement (Dubben et Beck-Bornholdt 2005). Ceci entraîne une surreprésentation des études primaires démontrant des effets positifs et peut affecter les conclusions de la revue. Il est recommandé de diversifier les sources de données et les types de documents à inclure. Par exemple, faire une recherche de la littérature grise pour inclure des rapports scientifiques et des thèses non publiées. En outre, il est possible qu'un biais de publication sélective des résultats existe (Martin, Renaud et Dagenais 2013), c'est-à-dire que les auteurs et autrices rapportent davantage les conséquences intentionnelles et attendues des interventions que les conséquences non intentionnelles et inattendues, ce qui est un défi dans notre exemple sur le FBR. Enfin, le biais relié à l'évaluateur ou à l'évaluatrice et à son interprétation des études peut être minimisé en impliquant au moins deux personnes dans l'évaluation et la synthèse des études. Dans une revue systématique mixte, il est recommandé d'avoir une équipe avec des expertises complémentaires en méthodologies et méthodes de recherche qualitative et quantitative afin de faciliter la synthèse et le jugement sur la qualité des études (Pluye et Hong 2014, Pluye *et al.* 2016).

Les inconvénients d'une approche systématique sont principalement que cela requiert du temps et des ressources. Selon l'objectif de la revue, le nombre de personnes impliquées, le nombre d'études à analyser et la synthèse à effectuer, la durée d'une revue systématique peut varier entre 6 à 24 mois (Khangura *et al.* 2012). Ainsi, il est important de s'assurer de la disponibilité des ressources et de justifier le choix d'effectuer une revue systématique mixte, par exemple en voulant tirer profit de la

complémentarité des données qualitatives et quantitatives pour expliquer un phénomène complexe et multidimensionnel.

Les revues systématiques peuvent être effectuées avec des logiciels de base (tels que Word ou Excel) ou spécialisés selon les méthodes de synthèse choisies (tels que QDA Miner, NVivo, SPSS, RevMan). Également, plusieurs logiciels spécialisés pour la revue systématique ont été développés tels que DistillerSR, Covidence et EPPI Reviewer. Ces logiciels sont conçus pour faciliter la gestion des données et peuvent être utilisés aux différentes étapes de la revue systématique, de la sélection des articles à l'extraction et l'analyse des données. Outre certains logiciels offerts gratuitement en ligne (par exemple, QDA Miner Lite), des frais mensuels ou annuels sont généralement requis. Leur utilisation dépend donc des ressources disponibles ainsi que de la complexité et l'ampleur de la revue.

Réaliser une revue systématique n'est pas toujours faisable. Dans une revue pour une thèse de doctorat ou une revue rapide pour un-e gestionnaire, il est rare d'avoir suffisamment de ressources et de temps pour effectuer une revue systématique. En conséquence, un-e étudiant-e de 2e ou 3e cycles ou un-e professionnel-le peut commencer par effectuer une revue des revues existantes. S'il n'y a pas de revues, la recherche d'études pertinentes peut être effectuée avec l'aide d'un-e bibliothécaire dans une base de données bibliographique pour obtenir un échantillon d'études. Puis la sélection, l'évaluation de la qualité des études incluses et l'extraction des données sont effectuées par une seule chercheuse ou un seul chercheur. Suivre ces étapes permet de gagner énormément de temps par rapport à une recherche documentaire non structurée et permet de produire un chapitre de thèse ou un rapport convaincant (à savoir une revue non exhaustive, mais transparente et reproductible).

Une des difficultés rencontrées dans la revue systématique mixte est l'intégration, c'est-à-dire la combinaison ou connexion des composantes qualitatives et quantitatives. Par exemple, une revue pourrait s'intéresser à diverses questions telles que l'efficacité du FBR, son coût-efficacité et ses conséquences non intentionnelles. Ces questions mobilisent divers types d'études et peuvent faire l'objet de revues distinctes. Ce qui caractérise la revue mixte est que l'ensemble des résultats sont intégrés afin de fournir une compréhension plus approfondie du FBR et émettre des recommandations plus précises. Sans cette intégration, une revue pourrait ne pas être considérée comme étant mixte, mais plutôt comme incluant plusieurs revues

indépendantes. Il est donc essentiel de bien décrire la façon dont les données sont intégrées. Les devis de synthèse présentés offrent des solutions pratiques pour intégrer les données. Aussi, le chapitre 7 présente diverses stratégies d'intégration pour les études primaires. Ces stratégies pourront être explorées pour être appliquées dans le cadre des revues systématiques. Dans les articles qui rapportent des revues systématiques mixtes, la discussion doit au minimum récapituler comment les résultats issus de données qualitatives et quantitatives se comparent ou se complètent. Elle doit refléter la valeur ajoutée de cette intégration et ses limites.

La revue systématique mixte est une forme de synthèse des connaissances en constante évolution. Le nombre d'articles rapportant des revues systématiques mixtes croît de manière exponentielle depuis 2010. Les bénéfices d'une revue systématique mixte peuvent être considérables, en particulier pour répondre à des questions complexes, tirer profit de la complémentarité des données quantitatives et qualitatives, et avoir une compréhension approfondie et complète d'un phénomène. La recherche sur les revues systématiques mixtes est encore jeune et les prochaines années seront riches en développements conceptuels et méthodologiques.

Références clés

Heyvaert, M., Hannes, K. et Onghena, P. (2016). *Using mixed methods research synthesis for literature reviews: The mixed methods research synthesis approach* (vol. 4). Thousand Oaks, CA: Sage Publications.

Il s'agit d'un livre méthodologique sur la revue mixte rédigé par des chercheuses et chercheurs en sciences sociales, psychologie et pédagogie de l'Université de Louvain en Belgique. Les différentes étapes pour mener une revue mixte y sont détaillées en débutant par la rédaction d'un protocole puis la recherche de documents pertinents, l'évaluation de la qualité des études, l'extraction et la synthèse des données et la publication d'une revue mixte.

Martin, V., Renaud, J. et Dagenais, P. (2013). *Les normes de production des revues systématiques : guide méthodologique.* https://www.inesss.qc.ca/fileadmin/doc/INESSS/DocuMetho/INESSS_Normes_production_revues_systematiques.pdf

Il s'agit d'un guide méthodologique rédigé en français qui propose

33 normes pour la production de revues systématiques. Ce guide ne porte pas spécifiquement sur la revue mixte, mais offre une excellente description des étapes à suivre pour effectuer une revue systématique. Il a été produit par des chercheuses et chercheurs de l'Institut national d'excellence en santé et en services sociaux (INESSS) *qui est une organisation francophone ayant comme mission de promouvoir l'excellence clinique et l'utilisation efficace des ressources dans le secteur de la santé et des services sociaux.*

Suri, H. (2014). *Towards methodologically inclusive research synthesis*. Abingdon, UK : Routledge.

Il s'agit d'un livre méthodologique sur les revues mixtes rédigé par une chercheuse en éducation à l'Université de Deakinen (Australie). Ce livre compte 11 chapitres dont les premiers portent sur les avancées méthodologiques sur la synthèse, les principes de la synthèse et les orientations épistémologiques. Les derniers chapitres expliquent comment mener une revue mixte, soit la formulation d'un objectif, la recherche documentaire, l'évaluation, la synthèse et l'interprétation des données.

Collaboration Cochrane : http://www.cochrane.org/

La Collaboration Cochrane est une organisation internationale à but non lucratif visant à promouvoir la pratique fondée sur des preuves scientifiques en santé. Elle regroupe plus de 30000 collaborateurs dans plus de 120 pays. Elle offre un soutien méthodologique et des formations sur la revue systématique. Il existe plusieurs branches francophones, dont Cochrane Canada – Francophone (http://ccf.cochrane.org/), Cochrane Suisse (http://swiss.cochrane.org/fr) et Cochrane France (http://france.cochrane.org). Aussi, un ouvrage méthodologique de référence sur les revues systématiques est disponible gratuitement : Cochrane Handbook for Systematic Reviews of Interventions : http://handbook.cochrane.org/

Collaboration Campbell : http://www.campbellcollaboration.org/

La Collaboration Campbell est une organisation similaire à la Collaboration Cochrane se spécialisant dans les sciences sociales, plus précisément dans les domaines de la justice, de l'éducation, du développement international et de la protection sociale).

Toolkit for Mixed Studies Reviews: http://toolkit4mixedstudiesreviews.pbworks.com

Ce site Internet sur la revue mixte est produit par des chercheuses et

chercheurs de l'Université McGill au Canada. Il se veut un site Internet collaboratif où toute personne intéressée par cette méthode est invitée à participer à sa mise à jour. Les étapes de la revue systématique mixte sont décrites en profondeur.

Références

Bai, A., Shukla, V. K., Bak, G. et Wells, G. (2012). *Quality assessment tools project report*. Ottawa : Canadian Agency for Drugs and Technologies in Health.

Barnett-Page, E. et Thomas, J. (2009). Methods for the synthesis of qualitative research: A critical review. *BMC Medical Research Methodology*, 9(1), 59.
doi: 10.1186/1471-2288-9-59

Bramer, W.M., Giustini, D., de Jonge, G. B., Holland, L. et Bekhuis, T. (2016). De-duplication of database search results for systematic reviews in EndNote. *Journal of the Medical Library Association : JMLA*, 104(3), 240-243.
doi: 10.3163/1536-5050.104.3.014

Busse, R., Orvain, J., Velasco, M., Perleth, M., Drummond, M., Jørgensen, T., … et Wild, C. (2002). Best practice in undertaking and reporting health technology assessments. *International Journal of Technology Assessment in Health Care*, 18(2), 361-422.
doi: 10.1017/S0266462302000284

Carroll, C. et Booth, A. (2015). Quality assessment of qualitative evidence for systematic review and synthesis: Is it meaningful, and if so, how should it be performed? *Research Synthesis Methods*, 6(2), 149-154.
doi: 10.1002/jrsm.1128

Chalmers, I., Hedges, L. V. et Cooper, H. (2002). A brief history of research synthesis. *Evaluation and the Health Professions*, 25(1), 12-37.
doi: 10.1177/0163278702025001003

Cochrane Effective Practice and Organisation of Care. 2012. LMIC Filters.
http://epoc.cochrane.org/lmic-filters

Cooper, H.M. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in Society*, 1(1), 104-126.

Creswell, J. W., Klassen, A. C., Clark, V. L. P. et Smith, K. C. (2011). *Best Practices for Mixed Methods Research in the Health Sciences*. Site du National Institutes of Health, Office of Behavioral and Social Sciences Research. http://obssr.od.nih.gov/mixed_methods_research

Crowe, M. et Sheppard, L. (2011). A review of critical appraisal tools show they lack rigor: alternative tool structure is proposed. *Journal of Clinical Epidemiology*, 64(1), 79-89. doi: 10.1016/j.jclinepi.2010.02.008

Dixon-Woods, M., Agarwal, S., Jones, D., Young, B. et Sutton, A. (2005). Synthesising qualitative and quantitative evidence: A review of possible methods. *Journal of Health Services Research and Policy*, 10(1), 45 – 53. doi: 10.1258/1355819052801804

Dubben, H.-H. et Beck-Bornholdt, H.-P. (2005). Systematic review of publication bias in studies on publication bias. *British Medical Journal*, 331(7514), 433-434.

El Sherif, R., Pluye, P., Gore, G., Granikov, V. et Hong, Q. N. (2016). Performance of a mixed filter to identify relevant studies for mixed studies reviews. *Journal of the Medical Library Association : JMLA*, 104(1), 47-51.

Fritsche, G.B., Soeters, R. et Meessen, B. (2014). *Boîte à outils : financement basé sur la performance*. Montreuil, FRA : Pearson France.

Gedda, M. (2015). Traduction française des lignes directrices PRISMA pour l'écriture et la lecture des revues systématiques et des méta-analyses. *Kinésithérapie, la Revue*, 15 (157), 39-44.

Gorter, A. C. et Meessen, B. (2013). *Evidence review, results-based financing of maternal and newborn health care in low-and lower-middle-income countries*. Germany: Study commissioned and funded by the German Federal Ministry for Economic Cooperation and Development through the sector project PROFILE at Deutsche Gesellschaft für Internationale Zusammenarbeit, Bonn et Eschborn.

Grant, M.J. et Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information and Librairies Journal*, 26(2), 91-108. doi: 10.1111/j.1471-1842.2009.00848.x

Harden, A. (2010). Mixed-methods systematic reviews: Integrating quantitative and qualitative findings. *Focus*, (25), 1-7.

Heyvaert, M., Hannes, K., Maes, B. et Onghena, P. (2013). Critical appraisal of mixed methods studies. *Journal of Mixed Methods Research*, 7(4), 302-327. doi: 10.1177/1558689813479449

Hong, Q. N., Pluye, P., Bujold, M. et Wassef, M. (2017). Convergent and sequential synthesis designs: Implications for conducting and reporting systematic reviews of qualitative and quantitative evidence. *Systematic Reviews*, 6(1), 61. doi: 10.1186/s13643-017-0454-2

Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C. et Vedel, I. 2018. « Mixed Methods Appraisal Tool (MMAT), version 2018 ». Registration of Copyright (#1148552), Canadian Intellectual Property Office, Industry Canada.

Khangura, S., Konnyu, K., Cushman, R., Grimshaw, J. et Moher, D. (2012). Evidence summaries: The evolution of a rapid review approach. *Systematic Reviews*, 1(1), 1-9. doi:10.1186/2046-4053-1-10

Langlois, A., Nie, J. Y., Thomas, J., Hong, Q. N. et Pluye, P. 2018. « Discriminating Between Empirical Studies and Nonempirical Works Using Automated Text Classification ». *Research Synthesis Methods*. doi: 10.1002/jrsm.1317

Martin, V., Renaud, J. et Dagenais, P. (2013). *Les normes de production des revues systématiques : guide méthodologique*. Montréal, QC : Institut national d'excellence en santé et en services sociaux (INESSS).

Mays, N., Pope, C. et Popay, J. (2005). Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field. *Journal of Health Services Research and Policy*, 10(Suppl. 1), 6 – 20. doi: 10.1258/1355819054308576

Moher, D., Liberati, A., Tetzlaff, J. et Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Public Library of Science : Medicine*, 6(7), :e1000097. doi: 10.1371/journal.pmed.1000097

Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage Publications.

Pace, R., Pluye, P., Bartlett, G., Macaulay, A. C., Salsberg, J., Jagosh, J. et Seller, R. (2012). Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review. *International Journal of Nursing Studies*, 49(1), 47-53.
doi: 10.1016/j.ijnurstu.2011.07.002

Paré, G., Trudel, M.-C., Jaana, M. et Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information et Management*, 52(2), 183-199.

Clark, V. L. P. et Ivankova, N. V (dir.). (2015). *Mixed methods research: A guide to the field* (vol. 3). Thousand Oaks, CA: Sage Publications.

Pluye, P., Gagnon, M. P., Griffiths, F. et Johnson-Lafleur, J. (2009). A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in mixed studies reviews. *International Journal of Nursing Studies*, 46(4), 529-46.
doi: 10.1016/j.ijnurstu.2009.01.009

Pluye, P. et Hong, Q. N. (2014). Combining the power of stories and the power of numbers: Mixed methods research and mixed studies reviews. *Annual Review of Public Health*, 35, 29-45.
doi:10.1146/annurev-publhealth-032013-182440

Pluye, P., Hong, Q. N., Bush, P. L. et Vedel, I. (2016). Opening-up the definition of systematic literature review: The plurality of worldviews, methodologies and methods for reviews and syntheses. *Journal of Clinical Epidemiology*, 73, 2-5.
doi:dx.doi.org/10.1016/j.jclinepi.2015.08.033

Pluye, P., Robert, E., Cargo, M., Bartlett, G., O'Cathain, A., Griffiths, F., ... et Rousseau, M.C. (2011). Proposal: A mixed methods appraisal tool for systematic mixed studies reviews. *Montréal : McGill University*, 1-8. http://mixedmethodsappraisaltoolpublic.pbworks.com/

Popay, J (dir.). (2006). *Moving beyond effectiveness in evidence synthesis: Methodological issues in the synthesis of diverse sources of evidence*. Londres, UK : National Institute for Health and Care Excellence (NICE).

Rathbone, J., Carter, M., Hoffmann, T. et Glasziou, P. (2015). Better duplicate detection for systematic reviewers: Evaluation of Systematic Review Assistant-Deduplication Module. *Systematic Reviews*, 4(1), 6.

Rogers, E. M. (2003). *Diffusion of Innovations* (5e éd.). New York : Free Press.

Souto, R. Q., Khanassov, V., Hong, Q. N., Bush, P. L., Vedel, I. et Pluye, P. (2015). Systematic mixed studies reviews: Updating results on the reliability and efficiency of the Mixed Methods Appraisal Tool. *International Journal of Nursing Studies*, 52(1), 500-501. doi:10.1016/j.ijnurstu.2014.08.010

Sutton, A.J. et Higgins, J. (2008). Recent developments in meta-analysis. *Statistics in Medicine*, 27(5), 625-650. doi: 10.1002/sim.2934

Thomas, J. et Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, 8, 45. doi: 10.1186/1471-2288-8-45

Tricco, A. C., Soobiah, C., Antony, J., Cogo, E., MacDonald, H., Lillie, E., … et Perrier, L. (2016). A scoping review identifies multiple emerging knowledge synthesis methods, but few studies operationalize the method. *Journal of Clinical Epidemiology*, 73, 19-28. doi: 10.1016/j.jclinepi.2015.08.030

Whittemore, R. et Knafl, K. (2005). The integrative review: Updated methodology. *Journal of Advanced Nursing*, 52(5), 546-553. doi: 10.1111/j.1365-2648.2005.03621.x

Witter, S., Fretheim, A., Kessy, F. L. et Lindahl, A. K. (2012). Paying for performance to improve the delivery of health interventions in low-and middle-income countries. *Cochrane Database of Systematic Review*, (2), CD007899.

Résumé / Abstract / Resumen

Les revues systématiques mixtes sont des revues de littérature utilisant une approche systématique pour combiner des études quantitatives, qualitatives et mixtes. Elles sont de plus en plus populaires, car elles permettent une compréhension approfondie des phénomènes et interventions complexes. Dans ce chapitre, huit étapes de la revue systématique mixte sont présentées : (a) formulation d'une question de recherche, (b) définition des critères d'éligibilité, (c) identification des

sources documentaires, (d) (élaboration de stratégie de recherche documentaire, (e) sélection des études pertinentes, (f) évaluation de la qualité des études retenues, (g) extraction des données des études retenues et (h) synthèse des données extraites. Ces étapes sont illustrées à partir d'un exemple tiré d'une revue systématique mixte sur les conséquences non intentionnelles d'une intervention reposant sur le financement basé sur les résultats. De plus, afin de guider l'intégration des données quantitatives et qualitatives, quatre devis de synthèse sont présentés : (a) devis de synthèse convergent avec intégration au niveau des données, (b) devis de synthèse convergent avec intégration au niveau des résultats de synthèse, (c) devis de synthèse convergent avec intégration au niveau de l'interprétation et (d) devis de synthèse séquentiel. Enfin, quelques défis des revues systématiques mixtes sont présentés dans l'analyse réflexive.

***

Systematic mixed studies reviews are literature reviews using a systematic approach to combine quantitative, qualitative and mixed methods studies. They are growing in popularity owing to their potential to provide a deep understanding of complex health interventions and problems. In this chapter, eight steps of systematic mixed studies reviews are presented: (a) formulate a review question, (b) define eligibility criteria, (c) identify sources of information, (d) select relevant studies, (e) appraise the quality of studies, (f) extract data and (g) synthesize extracted data from included studies. These steps are illustrated using an example of systematic mixed studies review on unintended consequences of performance-based financing intervention. Also, to provide guidance on how to combine qualitative and quantitative evidence, four synthesis designs are presented: (a) data-based convergent synthesis design, (b) results-based convergent synthesis design, (c) parallel-results convergent synthesis design, and (d) sequential synthesis design. This chapter ends with some challenges that can be encountered when conducting a systematic mixed studies review.

***

Las revisiones sistemáticas de estudios mixtos son revisiones de la literatura que utilizan un enfoque sistemático para combinar estudios cuantitativos, cualitativos y de métodos mixtos. Su popularidad va en aumento debido a su potencial para proporcionar una comprensión profunda de las intervenciones y los problemas de salud complejos. En este

capítulo se presentan ocho pasos de revisiones sistemáticas de estudios mixtos: (a) formular una pregunta de revisión, (b) definir los criterios de elegibilidad, (c) identificar las fuentes de información, (d) seleccionar los estudios pertinentes, (e) evaluar la calidad de los estudios, (f) extraer los datos y (g) sintetizar los datos extraídos de los estudios incluidos. Estos pasos se ilustran con un ejemplo de revisión sistemática de estudios mixtos sobre las consecuencias no deseadas de la intervención financiera basada en el desempeño. Además, para proporcionar orientación sobre cómo combinar la evidencia cualitativa y cuantitativa, se presentan cuatro diseños de síntesis: a) diseño de síntesis convergente basado en datos, b) diseño de síntesis convergente basado en resultados, c) diseño de síntesis convergente basado en resultados paralelos, y d) diseño de síntesis secuencial. Este capítulo termina con algunos de los desafíos que se pueden encontrar al realizar una revisión sistemática de estudios mixtos.

***

Quan Nha Hong est candidate au doctorat au département de médecine de famille de l'Université McGill et boursière des Instituts de recherche en santé du Canada (IRSC). Elle est ergothérapeute de formation. Elle possède une maîtrise en sciences cliniques de l'Université de Sherbrooke et une maîtrise en évaluation et gestion des technologies de la santé de l'Université de Montréal. Elle s'intéresse aux revues systématiques, méthodes de synthèse, méthodes mixtes et évaluation de la qualité méthodologique des études. Son projet de doctorat porte sur la mise à jour d'un outil d'évaluation de la qualité des études quantitatives, qualitatives et aux méthodes mixtes, le *Mixed Methods Appraisal Tool* (MMAT).

Anne-Marie Turcotte-Tremblay est candidate au doctorat en santé publique avec une spécialisation en organisation des soins à l'École de santé publique de l'Université de Montréal (ESPUM). Elle est également affiliée à l'Institut de recherche en santé publique de l'Université de Montréal (IRSPUM). Sa thèse de doctorat porte sur les conséquences non intentionnelles du financement basé sur les résultats combiné à des mesures de protection sociale au Burkina Faso. Anne-Marie est boursière des Instituts de recherche en santé du Canada (IRSC). Elle a coordonné plusieurs recherches internationales en santé des populations. Anne-Marie s'intéresse à l'évaluation d'interventions visant à réduire les inégalités sociales de santé.

Pierre Pluye est professeur titulaire au Département de médecine de

famille de l'Université McGill, chercheur-boursier senior du Fonds de recherche du Québec Santé (FRQS), et directeur de la composante Développement méthodologique à l'Unité SOUTIEN SRAP du Québec. Il est membre associé de l'École des sciences de l'information de l'Université McGill. Il est spécialiste en médecine familiale et en santé publique. Il détient un doctorat en promotion de la santé (Université de Montréal). Il a obtenu des bourses de carrière des Instituts de recherche en santé du Canada (IRSC) et du FRQS. En 2017, il a été élu membre de l'Académie canadienne des sciences de la santé, et a reçu le prix du « Chercheur de l'année » du Collège des médecins de famille du Canada. Il possède une expertise en méthodes mixtes. Ses recherches étudient les effets des informations issues de ressources électroniques sur les professionnels, les patients et le public.

Citation

Quan Nha Hong, Anne-Marie Turcotte-Tremblay et Pierre Pluye. (2019). Les revues systématiques mixtes. Un exemple à propos du financement basé sur les résultats. In *Évaluation des interventions de santé mondiale. Méthodes avancées*. Sous la direction de Valéry Ridde et Christian Dagenais, pp. 157-186. Québec : Éditions science et bien commun et Marseille : IRD Éditions.

7. L'intégration en méthodes mixtes

Cadre conceptuel pour l'intégration des phases, résultats et données qualitatifs et quantitatifs

PIERRE PLUYE

L'intégration en méthodes mixtes

Définition de la méthode

Les méthodes mixtes intègrent des méthodes qualitatives et quantitatives. Elles permettent de répondre à des questions de recherche et d'évaluation complexes.

Force de la méthode

Nous proposons trois types d'intégration et neuf stratégies pratiques, et toutes les combinaisons possibles des stratégies pour intégrer les phases, résultats et données qualitatives et quantitatives.

Défis de la méthode

Notre proposition peut combler un besoin (a) des chercheuses et chercheurs pour mieux planifier, effectuer et publier des études utilisant les méthodes mixtes, et (b) des enseignant-e-s pour former les étudiant-e-s en méthodes mixtes.

Les méthodes mixtes combinent les méthodes qualitatives (QUAL) et quantitatives (QUAN) en évaluation de programme, en recherche primaire et en revue de littérature (Creswell et Plano Clark 2011; Johnson, Onewuegbuzie et Turner 2007; Pluye et Hong 2014; Tashakkori et Teddlie 2010). Elles sont de plus en plus souvent utilisées en sciences sociales et de la santé. Au fil des ans, plusieurs stratégies pour intégrer les phases, résultats et données QUAL et QUAN ont été proposées, mais rarement conceptualisées et jamais testées (Greene 2008).

Pour tout chercheur ou chercheuse utilisant les méthodes mixtes, un défi consiste à planifier, conduire et rapporter simplement et clairement en quoi consistent les stratégies d'analyse spécifiques en méthodes mixtes et les combinaisons de ces stratégies. Ce chapitre propose un cadre conceptuel des principales combinaisons de stratégies qui sont utilisées en recherche primaire utilisant des méthodes mixtes. Dans le chapitre suivant, nous présenterons une veille collaborative des stratégies d'analyse en méthodes

mixtes qui a permis de tester cette proposition et de fournir des illustrations fondées sur des études utilisant les méthodes mixtes en santé mondiale.

Sont des méthodes mixtes toute combinaison de méthodologies et méthodes (incluant des devis de recherche et des techniques de collecte et d'analyse de données QUAL et QUAN) lorsque ces combinaisons satisfont trois conditions (Creswell et Plano Clark 2011) : (a) au moins une méthode QUAL et une méthode QUAN sont combinées (par exemple l'ethnographie et l'expérimentation), (b) chaque méthode est utilisée de façon rigoureuse par rapport aux critères généralement admis dans la méthodologie ou la tradition de recherche invoquée et (c) la combinaison des méthodes est effectuée au minimum via un devis en méthodes mixtes (planifié ou émergent) et une intégration des phases, des résultats et des données QUAL et QUAN. À noter que les méthodes QUAL et QUAN peuvent aussi être combinées (mais pas nécessairement) au niveau de la collecte de données (instrumentation mixte), la revue de la littérature (revue mixte) et l'interprétation des sciences (épistémologie, ontologie, téléologie et méthodologie) par l'équipe de chercheuses et chercheurs.

Inversement, ne sont pas considérées comme des méthodes mixtes : (a) une méthode QUAN avec une collecte ou analyse d'informations qualitatives qui ne sont pas des « données de recherche », car elles ne réfèrent pas à une méthodologie et méthode QUAL, (b) une méthode QUAL avec une collecte ou analyse d'informations quantitatives qui ne sont pas des « données de recherche », car elles ne réfèrent pas à un devis et une méthode QUAN, (c) les combinaisons de méthodes QUAN, (d) les combinaisons de méthodes QUAL, et (e) les juxtapositions de méthodes QUAL et QUAN sans intégration en termes d'approches, de devis, de techniques, de phases, résultats et données QUAL et QUAN.

Dans ce chapitre, nous présenterons les principaux éléments des méthodes mixtes (contexte) qui fondent le cadre conceptuel et nous proposerons trois types d'intégration (plan), neuf stratégies spécifiques aux méthodes (application) et sept combinaisons de stratégies (analyse réflexive) pour intégrer les phases, résultats et données qualitatifs et quantitatifs[1].

1. L'acronyme MM sera souvent utilisé dans la suite du chapitre pour désigner les méthodes mixtes.

Dans le chapitre suivant, nous présenterons des exemples issus de recherches en santé mondiale.

Contexte : pluralisme, devis, données et qualité

Pluralisme des méthodes mixtes

Dès 2003, Teddlie et Tashakkori ont affirmé la coexistence en méthodes mixtes de différentes interprétations du monde et des sciences par les chercheuses et chercheurs (Teddlie et Tashakkor, 2003). Ce pluralisme est illustré par de nombreuses publications (Niglas 2010). Dans ce chapitre, les méthodes mixtes sont fondées sur l'interprétation des sciences par les chercheuses et chercheurs (de manière explicite ou tacite) en termes d'épistémologie, ontologie, téléologie et méthodologie (Gendron 2001 ; Ridde et Dagenais 2012). Lorsqu'une équipe inclut des chercheuses et chercheurs en méthodes mixtes dont les interprétations des sciences diffèrent, la combinaison des méthodes nécessite des discussions d'ordre épistémologique, ontologique, téléologique et méthodologique. Par exemple, l'équipe peut combiner de manière transparente des approches ou traditions scientifiques qui reconnaissent les méthodes QUAL et QUAN telles que le postpositivisme selon Campbell, le constructionnisme social selon Hacking, le pragmatisme, le réalisme critique et les théories critiques, entre autres (Campbell 1988; De Waal 2005; Hacking 1999; Sayer 2000; Tyson 2014).

Principaux devis en méthodes mixtes

Les chercheuses et chercheurs peuvent s'inspirer des recommandations existantes pour planifier des recherches utilisant les méthodes mixtes (Creswell et Plano Clark 2011). Une classification usuelle des méthodes mixtes est fondée sur deux types de devis de recherche : séquentiel et convergent. Il existe aussi trois principales variantes de ces devis de base : les devis multiphases, multiniveaux et multiphases/multiniveaux.

Les devis de recherche séquentiels utilisent une méthode QUAL suivie d'une méthode QUAN (par exemple les résultats QUAN sont mobilisés pour généraliser statistiquement certains résultats QUAL), ou une méthode QUAN

suivie d'une méthode QUAL (par exemple les résultats QUAL sont mobilisés pour interpréter certains résultats QUAN). Dans tout devis séquentiel, les résultats de la phase-1 informent la phase-2. Dans ce chapitre, nous considérons qu'un projet de recherche peut être conceptualisé comme un processus organisationnel (par exemple un projet collectif impliquant des chercheuses et chercheurs QUAL et QUAN). La littérature sur les organisations (en management et gestion de projet) fournit alors une définition utile pour mieux comprendre ces séquences. Inspirés par Van de Ven (1992), nous proposons qu'une séquence soit définie (a) dans une perspective de développement (la phase-2 est construite sur les résultats de la phase-1), (b) comme un changement d'orientation du projet dans le temps (les résultats d'une première collecte/analyse de données [phase-1] informent une deuxième collecte/analyse de données [phase-2] qui est différente de la première en termes de méthodologie et méthode) et (c) comme une transition cognitive des chercheuses et chercheurs au moment de passer dans une nouvelle phase (passage d'une méthodologie et méthodes QUAL à QUAN, ou d'une méthodologie et méthodes QUAN à QUAL) (Van de Ven 1992).

Quant à eux, les devis convergents combinent les méthodes QUAL et QUAN dès la planification de la recherche ou au moment de la collecte et de l'analyse des données QUAN et QUAL. Dans un devis convergent, les méthodes QUAL et QUAN sont souvent (mais pas forcément) concomitantes. Ici encore, la littérature sur les organisations et les processus de prise de décision collective (par exemple les décisions prises par une équipe de chercheuses et chercheurs utilisant les méthodes mixtes) fournit des définitions de la convergence. Inspirés par Langley et ses collaborateurs, nous proposons que la convergence soit définie comme un processus d'amélioration progressive, successive et constante des collectes et des analyses des données QUAL et QUAN (convergence des données) et des interprétations des résultats (convergence des résultats). Les chercheuses et chercheurs travaillent de manière prospective, non linéaire, guidés par une représentation cognitive des données supplémentaires, ou des bases de données, ou des analyses des données, ou des résultats à créer (Langley, Mintzberg, Pitcher, Posada et Saint-Macary 1995).

Les variantes de ces devis consistent simplement à multiplier les phases ou les niveaux de collecte/analyse des données. D'une part, le devis multiphases inclut trois phases séquentielles (par exemple QUAL puis QUAN

puis QUAL) ou plus (Lisle 2013). La faisabilité de ce devis peut être problématique, car chaque phase dépend de l'achèvement en temps voulu de la précédente. D'autre part, le devis multiniveaux inclut deux niveaux d'analyse (par exemple QUAN au niveau individuel et QUAL au niveau organisationnel) ou plus (Dagenais, Nault-Brière, Dupont et Dutil 2008). Ce devis est fondé sur la convergence des résultats des analyses effectuées à chaque niveau. Youngs et Piggot-Irvine ont combiné les deux variantes et utilisé un devis multiphases/multiniveaux (Youngs et Piggot-Irvine 2011).

Dans un devis MM (séquentiel, convergent, multiphases, multiniveaux, ou multiphases/multiniveaux), les devis QUAL les plus fréquemment combinés avec un devis QUAN sont la recherche qualitative descriptive ou interprétative, l'étude de cas exploratoire, l'ethnographie, la théorisation ancrée, la phénoménographie et les récits de vie ou biographies (Schwandt 2007). Dans un devis MM, les devis QUAN les plus fréquemment combinés avec un devis QUAL sont quant à eux les devis descriptifs QUAN (par exemple les enquêtes de prévalence ou d'incidence), les études non randomisées (par exemple l'enquête transversale analytique, la cohorte, le cas-comparaison ou la quasi-expérimentation) et les expérimentations randomisées contrôlées ou pragmatiques (Porta 2008). Une mention particulière doit être faite pour les devis « étude de cas » et « théorisation ancrée », car ceux-ci peuvent être QUAL et QUAN ; en conséquence, chacun d'eux peut suffire (n'a pas nécessairement besoin d'être combiné avec un autre devis) pour planifier un devis de méthodes mixtes.

Techniques de collecte de données QUAL et QUAN en méthodes mixtes

En accord avec ces devis, toute combinaison de techniques de collecte de données QUAL et QUAN est possible en méthodes mixtes. Les chercheuses et chercheurs en MM sont créatifs et utilisent toutes les formes de collectes de données expérimentales, observationnelles ou simulées. Néanmoins, les techniques de collecte traditionnelles prédominent. Le questionnaire structuré tend à prédominer du côté des techniques QUAN, et l'entrevue ouverte ou semi-structurée tend à prédominer du côté des techniques QUAL (Bryman 2006). Le questionnaire structuré combinant des questions fermées (mesure standard testée en termes de validité et fidélité) et des questions ouvertes mérite une mention spéciale, car il constitue une source de questions récurrentes dans les cours et formations en MM. Ces

questions ouvertes peuvent faire partie des méthodes QUAL ou QUAN selon la façon dont elles sont conçues et utilisées. Par exemple, les réponses à des questions ouvertes constituent des données QUAL lorsqu'elles sont obtenues via une méthodologie et un processus de recherche scientifique (explicite, transparent et reproductible) QUAL rigoureux qui produit des résultats QUAL vraisemblables (crédibles, contextuels, confirmables, transférables). Les chercheuses et chercheurs connaissent les participant-e-s et interagissent avec eux (en reformulant les réponses ou en stimulant le développement des réponses) pour en savoir plus sur le contexte et mieux comprendre le sens des mots et des évènements.

En revanche, une réponse écrite dans une boîte de commentaire facultatif à une question ouverte qui est posée en fin de questionnaire structuré auto-administré anonyme en ligne ne peut pas être considérée comme une donnée QUAL. Dans les enquêtes épidémiologiques, les mots obtenus de cette manière fournissent traditionnellement quelques illustrations pour discuter les résultats QUAN. Ces mots constituent de l'information, mais ne constituent pas des données QUAL parce qu'ils ne peuvent pas être considérés comme des données obtenues via une méthodologie et un processus de recherche scientifique QUAL rigoureux, et ils ne peuvent pas servir à produire des résultats QUAL vraisemblables. Les chercheuses et chercheurs ne peuvent pas savoir qui a écrit ces mots et pourquoi (anonymisé) ; ils et elles ne peuvent pas interagir avec ceux et celles qui ont écrit et qui n'ont pas écrit (pour en savoir plus sur le contexte et mieux comprendre le sens des mots écrits ou les raisons pour lesquelles rien n'a été écrit). Dans le chapitre suivant, nous montrerons qu'environ 20 % des études en santé sont appelées « MM », mais ne satisfont pas les trois conditions de la définition, car elles utilisent des informations QUAL (ou QUAN) qui ne peuvent pas être considérées comme des données obtenues via une méthodologie et un processus de recherche scientifique QUAL (ou QUAN) rigoureux.

Évaluation de la qualité des méthodes mixtes

Ceci nous amène tout naturellement à l'évaluation de la qualité des méthodes mixtes. En combinant une revue de la littérature et une étude pilote avec des ateliers et des consultations avec des experts (Pace *et al.* 2012; Pluye, Gagnon, Griffiths et Johnson-Lafleur 2009), nous avons conçu

un outil d'évaluation de la qualité des études utilisant des méthodes QUAL, QUAN et MM : le MMAT (*Mixed Methods Appraisal Tool*) (Pluye *et al.* 2011). Le MMAT utilise différents critères pour différentes méthodes. Ainsi, il permet (a) d'évaluer la qualité méthodologique des composantes QUAL des MM à partir de critères utilisés pour évaluer les méthodes QUAL, (b) d'évaluer les composantes QUAN des MM à partir de critères utilisés en épidémiologie pour évaluer les méthodes QUAN ; et (c) d'évaluer la qualité générale des MM en termes de justification du devis, d'intégration des méthodes QUAL et QUAN, et de valeur ajoutée et limites de cette intégration. Le MMAT est disponible gratuitement sur internet (mixedmethodsappraisaltoolpublic.pbworks.com). Cet outil est populaire et tient sur une seule page, il est fourni avec un manuel qui en facilite l'utilisation. Des travaux récents ont testé la fidélité du MMAT et montré que cet outil est validé et efficient (Souto *et al.* 2015).

État des connaissances sur l'intégration en méthodes mixtes

Les analyses en MM « consistent en techniques d'analyses appliquées aux données QUAL et QUAN » (procédures standards utilisées en épidémiologie/statistiques et en recherche qualitative), et en techniques qui permettent d'intégrer des phases, résultats et données QUAL et QUAN « à un seul moment du processus de recherche utilisant des MM ou à plusieurs moments » (Creswell et Plano Clark 2011, p. 212). Ces dernières sont regroupées en stratégies spécifiques aux MM qui seront présentées dans la section « Application » ci-dessous, sous forme de proposition. Dans la section « Analyse », nous présenterons les combinaisons de ces stratégies.

Cette proposition est fondée sur une synthèse des principales publications méthodologiques de référence. Cette synthèse est innovante pour deux raisons : elle intègre toutes les stratégies spécifiques connues et toutes les combinaisons de stratégies possibles. De plus, cette proposition est associée avec un système de veille qui permet de la tester et d'identifier des stratégies émergentes (imprévues).

Un portrait général des analyses en MM selon 13 dimensions éclectiques a été récemment publié (Onwuegbuzie et Combs 2010) et inclue la séquence temporelle des analyses, la priorité donnée à telle ou telle analyse, l'orientation des analyses (cas, variables, processus), l'interdépendance des

analyses, les liens entre les analyses et les autres aspects des devis (philosophie, objectifs et formes de données) et le mode de généralisation (statistique ou théorique) des résultats. En dehors de ce portrait général, les publications méthodologiques en MM et les livres de référence proposent des stratégies de base non combinées (chaque stratégie étant présentée comme un processus nécessaire et suffisant pour obtenir des résultats) et les publications sont habituellement prescriptives (non testées avec des méthodes scientifiques comme les collectes et analyses de données de recherche) et limitées à quelques devis ou stratégies.

Pour planifier notre proposition, nous avons regroupé ces stratégies de base non combinées en trois grandes catégories de stratégies spécifiques : celles qui permettent (a) de connecter les phases QUAL et QUAN, (b) de comparer les résultats d'analyses QUAL et QUAN, et (c) d'assimiler les données QUAL et QUAN. Dans la section suivante, nous allons définir chacune de ces catégories successivement. Ces catégories ont trois avantages : elles ont été clairement définies en appliquant les principes d'harmonisation des termes utilisés en documentation (sciences de l'information) (International Standards Organization [ISO] 2009; Roche 2012), elles intègrent les terminologies antérieures (Bazeley 2009; Greene 2007; Teddlie et Tashakkori 2009), et elles réfèrent aux trois types usuels d'intégration des méthodes QUAL et QUAN (connexion de phases, comparaison de résultats et assimilation des données).

La connexion de phases est appelée « développement séquentiel » (Bazeley), ou « corrélation et comparaison » (Greene) ou encore « analyse mixte séquentielle » (Teddlie et Tashakkori). La comparaison de résultats est appelée « triangulation et expansion » (Bazeley) ou « analyse inférentielle conjointe » (Greene) ou encore « analyse mixte parallèle » (Teddlie et Tashakkori). L'assimilation des données est appelée « transformation » (Bazeley), « transformation et consolidation » (Greene) ou « analyse mixte par conversion » (Teddlie et Tashakkori). De plus, nous avons repris deux des trois catégories proposées par Creswell, mais sans limiter chaque catégorie à un devis en MM. En effet, Creswell et ses collaborateurs limitent chaque catégorie à un devis : connexion de phases QUAL et QUAN (devis séquentiel), comparaison de résultats QUAL et QUAN, et transformation de données QUAL ou QUAN (devis convergent) (Creswell et Plano Clark 2011; Guetterman, Fetters et Creswell 2015).

Trois types d'intégration en MM

Greene, Teddlie et Tashakkori (Greene 2007; Teddlie et Tashakkori 2003, 2009, 2010) suggèrent trois principes des MM (complémentarité, tension dialectique, et unification) qui justifient conceptuellement les trois types d'intégration proposés (connexion de phases, comparaison de résultats, et assimilation des données). Ces trois types d'intégration ne sont pas mutuellement exclusifs (ils peuvent être combinés) et ne sont pas associés à un jugement de valeur (par exemple la connexion de phase ne constitue pas une meilleure ou moins bonne intégration que l'assimilation des données).

- Connexion de phases. Le principe de complémentarité est issu de la littérature suggérant que les interprétations des sciences (épistémologie, ontologie, méthodologie et téléologie) associées aux méthodes QUAL sont différentes et séparées de celles qui sont associées aux méthodes QUAN. Les méthodes de collecte et d'analyse des données QUAL et QUAN sont séparées. Les méthodes et résultats QUAL et QUAN sont présentés séparément dans les publications, mais au moins une phrase suggère que les résultats QUAL et QUAN sont complémentaires. L'intégration (transition cognitive) survient lors de la connexion entre deux phases (par exemple entre une phase QUAL et une phase QUAN).
- Comparaison de résultats. Le principe de tension dialectique provient des publications suggérant que les interprétations des sciences associées aux méthodes QUAL et QUAN sont différentes et interdépendantes (leur juxtaposition générant des tensions créatrices de découverte et d'innovation). Les méthodes de collecte et d'analyse des données QUAL et QUAN sont séparées ou interconnectées à un ou plusieurs moments, et les résultats sont interconnectés via un processus de comparaison. Les similarités, différences et contradictions entre résultats QUAL et QUAN sont explicitées (guidées par une représentation cognitive des résultats à créer). Par exemple, des divergences entre les résultats QUAN et QUAL sont mentionnées dans les publications, et au moins une phrase suggère comment les résoudre.
- Assimilation des données. Le troisième principe, l'unification, est centré sur une question ou une théorie ou une interprétation des sciences associées aux méthodes QUAL et QUAN. Il correspond à deux courants de pensée : d'une part, les méthodes répondent à des questions

de recherche et mobilisent des théories qui unifient l'utilisation des MM sur une question ou une théorie (indépendamment des interprétations des sciences). D'autre part, plusieurs interprétations des sciences permettent directement l'utilisation des MM (unification sur une interprétation). Ce principe justifie l'assimilation des données (guidée par une représentation cognitive des résultats à créer). Les données QUAL et QUAN peuvent être transformées sous une seule forme QUAL (thèmes) ou QUAN (variables) ou fusionnées cas par cas.

Application

Cadre conceptuel

Nous proposons un cadre conceptuel incluant ces trois types d'intégration, neuf stratégies spécifiques, et sept combinaisons de stratégies, pour intégrer les phases, résultats et données QUAL et QUAN en MM. Cette conceptualisation est résumée dans le tableau 1 et la figure 1. Elle suggère que toute étude utilisant des MM peut combiner plusieurs stratégies. Elle est inspirée du constructionnisme social (Hacking 1999) : toute stratégie produit un objet mixte (*mixed kind*) par effet de boucle (*looping effect*) entre les phases, les résultats ou les données QUAL et QUAN. Autrement dit, chaque stratégie est définie par une ressource (un intrant dans le processus de gestion du projet de recherche utilisant les MM), un processus (effet de boucle) et un produit (objet mixte). Toute combinaison de stratégies multiplie les effets de boucle et les objets mixtes.

Pour chaque type d'intégration, les stratégies spécifiques que nous proposons (sans limiter chaque stratégie à un seul devis) correspondent aux stratégies de base décrites par des auteurs et autrices souvent citées en méthodes mixtes. En évaluation, quatre stratégies de base sont habituellement décrites : développement de typologie ou taxonomie, analyse de cas extrêmes, transformation de données et consolidation ou fusion de données (Caracelli et Greene 1993). En sciences de la santé, les stratégies de base habituellement décrites sont la comparaison de résultats (triangulation), le suivi d'un filon ou d'un fil d'Ariane (*follow a thread*), la fusion des données QUAL et QUAN pour chaque cas (méta-matrice), et la transformation des données QUAL en QUAN (quantification) ou de données

QUAN en QUAL (qualitativation) (O'Cathain, Murphy, et Nicholl 2010; Sandelowski 2000). Les neuf stratégies d'intégration spécifiques aux MM sont présentées ci-dessous.

Premier type d'intégration : connexion de phases (3 stratégies spécifiques)

Stratégie 1.1. Connexion d'une phase QUAL à une phase QUAN

Dans cette stratégie, la ressource est constituée des résultats de la phase-1 QUAL (Creswell et Plano Clark 2011). Le processus est la connexion des résultats QUAL avec la collecte et l'analyse des données de la phase-2 QUAN; par exemple, les chercheuses et chercheurs utilisent des résultats de la phase-1 QUAL pour construire le questionnaire de la phase-2 QUAN. Le produit mixte consiste en une collecte/analyse de données QUAN qui est informée (structurée) par des résultats QUAL. Cette stratégie permet d'atteindre les objectifs suivants (entre autres) : (a) développer une typologie (phase-1 QUAL) et une taxonomie (phase-2 QUAN) ; (b) créer un instrument (phase-1 QUAL) et l'utiliser pour une mesure (phase-2 QUAN) ; (c) valider un instrument (validation de contenu QUAL puis validation de construit QUAN ; ou validation écologique de contenu QUAL -représentativité des questions- puis QUAN -pertinence des questions-) ; (d) concevoir une intervention (phase-1 QUAL) et en évaluer la mise en œuvre (phase-2 QUAN) ; et (e) construire un cadre conceptuel ou théorique (phase-1 QUAL) et le tester (phase-2 QUAN).

Stratégie 1.2. Connexion d'une phase QUAN à une phase QUAL

La ressource est constituée des résultats de la phase-1 QUAN (Creswell et Plano Clark, 2011). Le processus est la connexion des résultats QUAN avec la collecte et l'analyse des données de la phase-2 QUAL; par exemple, les chercheuses et chercheurs utilisent des résultats de la phase-1 QUAN pour identifier des catégories d'informateurs clés de la phase-2 QUAL. Le produit mixte consiste en une collecte/analyse de données QUAL qui est informée par des résultats QUAN. Cette stratégie permet d'atteindre les objectifs suivants (entre autres) : (a) valider un instrument (les résultats QUAL peuvent

explorer des explications des résultats QUAN) ; (b) expliquer des résultats QUAN en utilisant des résultats QUAL obtenus avec un échantillon raisonné d'informateurs clés, p. ex., pour mieux comprendre les différences entre des groupes (intervention vs. contrôle ; favorisé vs. défavorisé) ; et (c) contribuer à expliquer des cas extrêmes.

Cas particulier de 1.1 et 1.2. Suivi d'un fil d'Ariane entre phases QUAL et QUAN

Ce cas particulier est appelé stratégie du fil d'Ariane (*follow the thread*) (O'Cathain *et al.* 2010). La ressource est constituée des résultats d'une phase QUAL ou QUAN. Le processus est constitué de la connexion des résultats de cette phase avec la collecte et l'analyse des données de la phase suivante (QUAN ou QUAL), de l'identification d'un thème QUAL ou d'une variable clé QUAN lors de ces phase, et de la re-analyse en profondeur des données liées à ce thème ou cette variable de manière prospective et rétrospective (métaphoriquement comme pour dérouler un fil d'Ariane à l'aller dans une grotte, puis pour suivre ce fil au retour à l'air libre). Le produit mixte consiste en une collecte/analyse de données QUAL ou QUAN qui est informée par des résultats QUAN ou QUAL.

Deuxième type d'intégration : comparaison de résultats (3 stratégies spécifiques)

Stratégie 2.1. Comparaison des résultats obtenus de manière séparée

Dans cette stratégie, la ressource est constituée des résultats QUAL et QUAN obtenus via des collectes et analyses de données séparées (Creswell et Plano Clark 2011). Le processus est une comparaison par les chercheuses et chercheurs des similarités et différences entre résultats QUAL et QUAN, par exemple en les juxtaposant dans un tableau dont chaque ligne correspond à un thème QUAL principal et à la (aux) variable(s) correspondante(s). Le produit mixte consiste en une interprétation ou une décision. Typiquement, un tableau de comparaison est formé de quatre colonnes : sous-titre (colonne 1), résultats QUAL (colonne 2), résultats QUAN (colonne 3) et interprétation ou décision (colonne 4).

Stratégie 2.2. Comparaison des résultats obtenus de manière interdépendante

La ressource est constituée des résultats QUAL et QUAN obtenus via des collectes et analyses de données interdépendantes (Creswell et Plano Clark 2011). Comme dans la stratégie précédente, le processus est une comparaison par les chercheuses et chercheurs des similarités et différences entre résultats QUAL et QUAN. Le produit mixte consiste en une interprétation ou une décision qui tient compte des interdépendances de manière explicite, par exemple lorsqu'une variable QUAN est dérivée directement d'un thème QUAL (ou les participant-e-s QUAL recrutés parmi les participant-e-s QUAN). Lorsqu'il y a interdépendance, les chercheuses et chercheurs s'attendent principalement à présenter des similarités, par exemple une corroboration des résultats QUAL par les résultats QUAN (ou vice-versa). Plusieurs hypothèses peuvent être formulées pour expliquer les différences s'il y a lieu (chaque différence pouvant constituer une menace à la validité interne des résultats QUAN ou à la crédibilité des résultats QUAL).

Cas particulier des stratégies 2.1 et 2.2. Divergence des résultats QUAL et QUAN

Dans ce cas particulier, les chercheuses et chercheurs comparent les résultats QUAL et QUAN en mettant l'accent sur les divergences (contradictions, discordances ou dissonances). Comme précédemment, le processus est une comparaison par les chercheuses et chercheurs des résultats QUAL et QUAN. Le produit mixte consiste en une interprétation ou une décision pour chaque divergence. Dans la littérature en MM, nous avons trouvé des exemples qui illustrent deux principales stratégies que les chercheuses et chercheurs utilisent pour faire face aux divergences entre résultats QUAL et QUAN : réconciliation ou initiation d'un nouveau projet de recherche (Pluye, Grad, Levine, et Nicolau 2009).

Troisième type d'intégration : assimilation des données (3 stratégies spécifiques)

Stratégie 3.1. Transformation de données QUAL en données QUAN (quantification)

Dans cette stratégie, la ressource est constituée de toutes les données ou d'une partie des données QUAL d'une recherche qualitative, et de données QUAN. Le processus est la transformation ou conversion des données QUAL en données QUAN de manière à ce que les données QUAL transformées puissent être analysées avec les données QUAN. Par exemple, les chercheuses et chercheurs utilisent la technique séculaire d'analyse de contenu quantitative (Neuendorf, 2002) : ils et elles transforment des thèmes clés en un petit nombre de variables, et les données QUAL correspondantes en valeurs numériques, en utilisant une grille et un manuel de codage valide et fidèle. Le produit mixte consiste en données QUAN (combinant des données QUAN et des données QUAL transformées en données QUAN) qui peuvent être analysées en utilisant les techniques statistiques habituelles. La rigueur de cette stratégie repose sur la validité et la fidélité inter-juge de l'analyse de contenu quantitative (la grille et le manuel de codage étant l'instrument de mesure). À notre avis, plusieurs chercheuses et chercheurs critiquent à tort cette stratégie sous prétexte que cela « réduirait » les données QUAL. Dans les faits, rien n'empêche que toutes les données QUAL soient analysées de manière QUAL en parallèle. La transformation de données QUAL en données QUAN produit une plus-value en termes d'analyses et de résultats supplémentaires.

Stratégie 3.2. Transformation de données QUAN en données QUAL (qualitativation)

La ressource est constituée de toutes les données ou une partie des données QUAN d'une recherche épidémiologique ou statistique, et de données QUAL. Le processus est la transformation ou conversion des données QUAN en données QUAL de manière à ce que les données QUAN transformées puissent être analysées avec les données QUAL. Les chercheuses et chercheurs utilisent une analyse interprétative (par exemple thématique) et transforment des résultats statistiques en histoires ou

narratifs (par exemple organisés par thème) (Sandelowski, 2000). Le produit mixte consiste en données QUAL (combinant des données QUAL et des données QUAN transformées en données QUAL) qui peuvent être analysées de manière narrative. Comme dans la stratégie précédente, rien n'empêche que toutes les données QUAN soient analysées de manière statistique en parallèle. La transformation de données QUAN en données QUAL fournit une plus-value (analyses et résultats supplémentaires).

Stratégie 3.3. Fusion des données QUAL et QUAN

Dans cette dernière stratégie, la ressource est constituée de cas (par exemple des personnes, des organisations ou des évènements, entre autres), et de données QUAL et QUAN (O'Cathain *et al.* 2010). Le processus est la fusion des données QUAL et QUAN au cas par cas (analyse intra-cas) de manière à permettre des analyses inter-cas. Le produit mixte consiste en un nouvel ensemble de données qui peuvent être analysées de manière statistique (par exemple la description des cas et tests d'hypothèses) ou narrative (par exemple la comparaison des histoires de cas). La rigueur de cette stratégie est fondée sur la définition des cas et une collecte de données QUAL et QUAN similaires pour chaque cas. Par exemple, des clinicien-ne-s ont évalué avec un questionnaire validé les effets cognitifs et l'utilisation des informations qu'ils et elles trouvent dans des ressources électroniques spécialisées (étude longitudinale QUAN); les cas ont été définis comme des recherches d'information où l'information clinique a été utilisée pour un-e patient-e spécifique. Pour chaque cas, les clinicien-ne-s ont été interviewés et ont décrit les effets de l'utilisation de ces informations sur la santé des patients (études de cas multiples QUAL), les données QUAN et QUAL ont été fusionnées en vignettes cliniques (chaque vignette décrivant un cas) et ces vignettes ont permis de construire une « méta-matrice mixte » et d'estimer la proportion de patients dont la santé bénéficie des recherches d'informations par les clinicien-ne-s (Pluye, Grad, Johnson-Lafleur *et al.* 2013 ; Pluye, Grad, Repchinsky *et al.* 2013).

Analyse réflexive : combinaisons de stratégies pour intégrer phases, résultats et données QUAL et QUAN

Une combinaison est définie par un ou plusieurs éléments d'un ensemble sans égard pour l'ordre de la sélection. Les sept combinaisons possibles des stratégies proposées sont présentées dans la Figure 1.

Ce chapitre a décrit le pluralisme des méthodes mixtes (MM) en termes d'interprétations des sciences (épistémologie, ontologie, méthodologie et téléologie) et a rappelé que les MM requièrent habituellement la collaboration d'experts en méthodes QUAL et en méthodes QUAN (par exemple un superviseur ou unesuperviseuse QUAL et un co-superviseur ou une co-superviseuse QUAN pour les doctorant-e-s) qui doivent réconcilier leur interprétation des sciences lorsque celles-ci diffèrent. De plus, ce chapitre a mentionné les principaux types de devis et critères de qualité en MM. À ces connaissances de base, nous ajoutons trois types d'intégration, neuf stratégies spécifiques aux MM et tout un nouveau monde à explorer avec des multiples possibilités de combinaisons de stratégies pour intégrer des phases, des résultats et des données QUAL et QUAN.

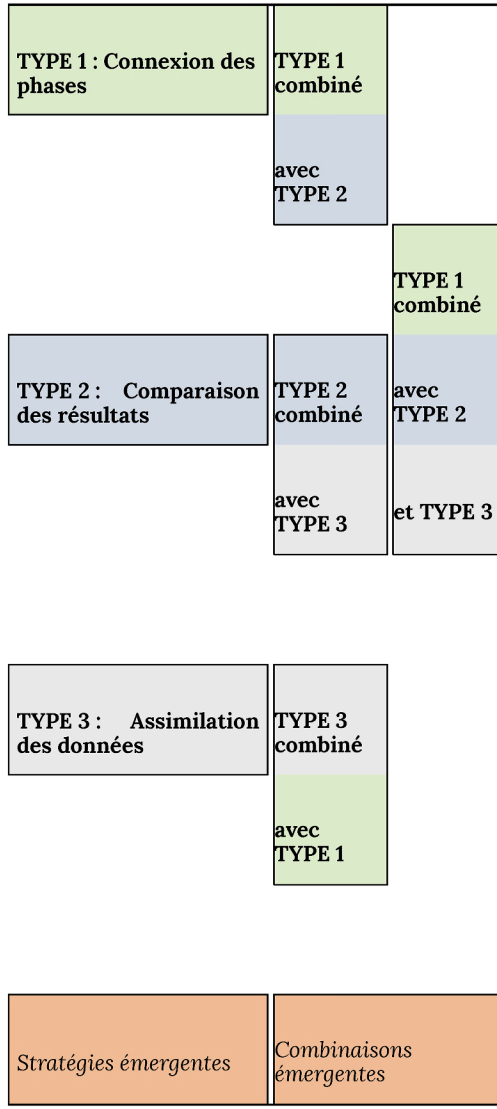
Ce chapitre a contribué aux connaissances en MM en termes de définition, pratique et innovation. Il a défini les concepts de séquence et convergence en lien avec la littérature sur la gestion de projets, en l'occurrence la gestion des projets de recherche ou d'évaluation utilisant les MM. Notre conceptualisation contribue aux connaissances sur les MM et se veut pratique. Elle peut guider les étudiant-e-s, les chercheuses et chercheurs et les évaluateurs et évaluatrices pour planifier, conduire et évaluer des études et des évaluations de programmes utilisant les MM, notamment en santé mondiale. Dans le chapitre suivant, nous illustrerons toutes ces stratégies et combinaisons avec des recherches en santé mondiale. Nous discuterons les forces et les faiblesses de nos propositions à la lumière de ces recherches.

Tableau 1. Neuf stratégies spécifiques pour intégrer des phases, résultats ou données qualitatives (QUAL) et quantitatives (QUAN) en méthodes mixtes (MM)

3 TYPES D'INTÉGRATION et 9 STRATÉGIES SPÉCIFIQUES	DÉFINITIONS
TYPE 1 : Connexion des phases	Connecter des phases QUAL et QUAN
Phase QUAL à phase QUAN	Connecter les résultats d'une phase-1 QUAL avec la collecte et l'analyse d'une phase-2 QUAN [1].
Phase QUAN à phase QUAL	Connecter les résultats d'une phase-1 QUAN avec la collecte et l'analyse d'une phase-2 QUAL [1].
Cas particulier de 1.1 et 1.2 : « Follow a thread »	Analyser les données QUAL (ou QUAN) et identifier les principaux thèmes (ou variables) qui nécessitent un approfondissement ; choisir un thème (ou variable) et réanalyser à travers les composantes QUAN (ou QUAL) [2].
TYPE 2 : Comparaison des résultats	Comparer des résultats QUAL et QUAN
2.1. Résultats QUAL et QUAN obtenus de manière séparée	Comparer les similarités et différences entre les résultats QUAL et QUAN issus de collectes et analyses de données séparées [1].
2.2. Résultats QUAL et QUAN obtenus de manière interdépendante	Comparer les similarités et différences entre les résultats QUAL et QUAN issus de collectes et analyses de données interdépendantes [1].
Cas particulier de 2.1 et 2.2 : Divergence des résultats QUAL et QUAN	Comparer en mettant l'accent sur les divergences (contradictions, discordances ou dissonances) entre les résultats QUAL et QUAN [3].
TYPE 3 : Assimilation des données	Assimiler des données QUAL et QUAN
3.1. Données QUAL en données QUAN	Assimiler les données QUAL et QUAN en transformant les données QUAL en données QUAN [1].
3.2. Données QUAN en données QUAL	Assimiler les données QUAL et QUAN en transformant les données QUAN en données QUAL [1].
3.3. Fusion des données QUAL et QUAN	Assimiler les données QUAL et QUAN en les fusionnant pour chaque cas dans une base de données additionnelle [1].
TYPE 4 : Stratégies émergentes	À découvrir en surveillant l'évolution des tendances en MM avec eSRAP

[1](Creswell et Plano Clark 2011); [2] (O'Cathain *et al.* 2010); [3](Pluye, Grad *et al.* 2009)

Figure 1. Sept combinaisons de stratégies pour intégrer des phases, résultats ou données qualitatives (QUAL) et Quantitatives (QUAN) en méthodes mixtes (MM)



Références clés

Méthodes mixtes francophonie :
http://methodesmixtesfrancophonie.pbworks.com

Cette nouvelle association internationale à but non lucratif vise à promouvoir les méthodes mixtes en français. En 2018, elle regroupe déjà 121 membres de 8 pays différents (Allemagne, Belgique, Canada, France, Rwanda, Sénégal et Suisse). Elle offre un soutien méthodologique, une bibliographie ouverte de publications des membres, et une liste d'experts, de ressources et de formations. Notamment, elle offre un accès gratuit au premier ouvrage collectif méthodologique de référence sur les méthodes mixtes en français (14 chapitres, 266 pages): Bujold, M., Hong, Q. N., Ridde, V., Bourque, C. J., Dogba, M. J., Vedel, I. et Pluye, P. (2018). *Oser les défis des méthodes mixtes en sciences sociales et sciences de la santé*. Montréal : 117e Cahiers scientifiques de l'ACFAS.

Creswell, J. W. et Clark, V. L. P. (2018). *Designing and conducting mixed methods research*(3e éd.). Thousand Oaks, CA: Sage Publications.

Il s'agit du livre méthodologique le plus populaire sur les méthodes mixtes. Les différentes étapes pour mener une recherche utilisant les méthodes mixtes y sont détaillées. Ce livre constitue un guide pratique pour les évaluateurs et évaluatrices, les étudiant-e-s et les chercheuses et chercheurs novices en méthodes mixtes. Il propose des devis de recherche en méthodes mixtes, des exemples de collecte et d'analyse des données, et des recommandations pour rédiger un protocole ou une publication.

Tashakkori, A. et Teddlie, C. (2010). *Handbook of mixed methods in social and behavioral research*. Thousand Oaks, CA: Sage Publications.

Ce livre collectif de référence inclut des chapitres sur quasiment tous les aspects importants des méthodes mixtes en sciences sociales et sciences de la santé. Il est rédigé par des chercheuses et chercheurs de plusieurs disciplines et pays qui sont reconnus dans leur domaine. Il présente un état des connaissances épistémologiques et méthodologiques sur les méthodes mixtes. C'est le compagnon des étudiant-e-s et des chercheuses et chercheurs qui utilisent les méthodes mixtes.

Références

Bazeley, P. (2009). Analyzing mixed methods data. Dans S. Andrew et E. Halcomb (dir.), *Mixed methods research for nursing and the health sciences* (p. x-x). Chichester : Wiley.

Bryman, A. (2006). Integrating quantitative and qualitative research: how is it done?. *Qualitative Research*, 6(1), 97-113.

Campbell, D. T. (1988). Qualitative knowing in action research. Dans S. Overman (dir.), *Methodology and epistemology for social science: selected papers of Donald T. Campbell* (p. 360-376). Chicago : The University of Chicago Press.

Caracelli, V. J. et Greene, J. C. (1993). Data analysis strategies for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 15(2), 195-207.

Creswell, J. W. et Clark, V. L. P. (2011). *Designing and conducting mixed methods research* (2e éd.). Thousand Oaks, CA: Sage Publications.

Dagenais, C., Nault-Brière, F., Dupont, D. et Dutil, J. (2008). Implementation and effects of a service coordination program for youths and their families in complex situations: A mixed evaluation design. *Children and Youth Services Review*, 30(8), 903-913.

De Waal, C. (2005). *On pragmatism*. Belmont : Wadsworth.

Gendron, S. (2001). *La pratique participative en santé publique : l'émergence d'un paradigme* (Thèse de doctorat). Université de Montréal, Montréal.

Greene, J. C. (2007). *Mixed methods in social inquiry* (édition, volume). San Francisco: Jossey Bass.

Greene, J. C. (2008). Is mixed methods social inquiry a distinctive methodology?. *Journal of Mixed Methods Research*, 2(1), 7-22. doi: 10.1177/1558689807309969

Guetterman, T. C., Fetters, M. D. et Creswell, J. W. (2015). Integrating quantitative and qualitative results in health science mixed methods research through joint displays. *The Annals of Family Medicine*, 13(6), 554-561.

Hacking, I. (1999). *The social construction of what?*. Cambridge, MA: Harvard University Press.

International Standards Organization (ISO). (2009). *Terminology work – Principles and methods*. http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=38109

Johnson, R. B., Onewuegbuzie, A. J. et Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2), 112-133.

Langley, A., Mintzberg, H., Pitcher, P., Posada, E. et Saint-Macary, J. (1995). Opening up decision making: the view from the black stool. *Organization Science*, 6(3), 260-279.

Lisle, J. D. (2013). Exploring the value of integrated findings in a multiphase mixed methods evaluation of the continuous assessment program in the Republic of Trinidad and Tobago. *International Journal of Multiple Research Approaches*, 7(1), 27-49.

Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage Publications.

Niglas, K. (2010). The multidimensional model of research methodology: an integrated set of continua. Dans A. Tashakkori et C. Teddlie (dir.), *Handbook of mixed methods in social and behavioral research* (2e éd., p. 215-236). Thousand Oaks, CA: Sage Publications.

O'Cathain, A, Murphy, E. et Nicholl, J. (2010). Three techniques for integrating data in mixed methods studies. *British Medical Journal*, 341, 1147-1150. doi: 10.1136/bmj.c4587

Onwuegbuzie, A. J. et Combs, J. P. (2010). Emergent data analysis techniques in mixed methods research: a synthesis. Dans A. Tashakkori et C. Teddlie (dir.), *Handbook of mixed methods in social and behavioral research* (p. 397-430). Thousand Oaks, CA: Sage Publications.

Pace, R., Pluye, P., Bartlett, G., Macaulay, A. C., Salsberg, J., Jagosh, J. et Seller, R. (2012). Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review. *International Journal of Nursing Studies*, 49(1), 47-53.

Pluye, P., Gagnon, M. P., Griffiths, F. et Johnson-Lafleur, J. (2009). A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies

in mixed studies reviews. *International Journal of Nursing Studies*, 46(4), 529-546.

Pluye, P., Grad, R. M., Johnson-Lafleur, J., Granikov, V., Shulha, M., Marlow, B. et Ricarte, I. (2013). The Number Needed to Benefit from Information (NNBI): proposal from a mixed methods research study with practicing family physicians. *The Annals of Family Medicine*, 11(6), 559-567. doi: 10.1370/afm.1565.

Pluye, P., Grad, R. M., Levine, A. et Nicolau, B. (2009). Understanding divergence of quantitative and qualitative data and results in mixed methods studies. *International Journal of Multiple Research Approaches*, 3(1), 58-72.

Pluye, P., Grad, R. M., Repchinsky, C., Jovaisas, B., Johnson-Lafleur, J., Carrier, M. E., Légaré, F. (2013). Four levels of outcomes of information-seeking: a mixed methods study in primary health care. *Journal of the American Society for Information Science and Technology* (JASIST), 64(1), 108–125.

Pluye, P. et Hong, Q. N. (2014). Combining the power of stories and the power of numbers: mixed methods research and mixed studies reviews. *Annual Review of Public Health*, 35, 29-45.

Pluye, P., Robert, E., Cargo, M., Bartlett, G., O'Cathain, A., Griffiths, F., Rousseau, M.C. (2011). *Proposal: a mixed methods appraisal tool for systematic mixed studies reviews*. http://mixedmethodsappraisaltoolpublic.pbworks.com

Porta, M. (2008). *A dictionary of epidemiology*. New York : Oxford University Press.

Ridde, V. et Dagenais, C. (2012). Introduction générale à l'évaluation de programme. Dans V. Ridde et C. Dagenais (dir.), *Approches et pratiques en évaluation de programme* (p. 11-29). Montréal : Presses de l'Université de Montréal.

Roche, C. (2012, mai). *Ontologie : entre terminologie et connaissance de spécialité*. Communication présentée au GLAT – Terminologies : textes, discours et accès aux savoirs spécialisés, Gênes. https://hal.archives-ouvertes.fr/hal-01180282/document

Sandelowski, M. (2000). Combining qualitative and quantitative sampling, data collection, and analysis techniques in mixed-method studies. *Research in Nursing et Health*, 23(3), 246-255.

Sayer, A. (2000). *Realism and social science*. Londres : Sage Publications.

Schwandt, T. A. (2007). *Dictionary of qualitative inquiry* (3e éd.). Thousand Oaks, CA: Sage Publications.

Souto, R. Q., Khanassov, V., Hong, Q. N., Bush, P. L., Vedel, I. et Pluye, P. (2015). Systematic mixed studies reviews: updating results on the reliability and efficiency of the mixed methods appraisal tool. *International Journal of Nursing Studies*, 52(1), 500-501.

Tashakkori, A. et Teddlie, C. (2010). *Handbook of mixed methods in social and behavioral research*. Thousand Oaks, CA: Sage Publications.

Teddlie, C. et Tashakkori, A. (2003). Major issues and controversies in the use of mixed methods in the social and behavioural sciences. Dans A. Tashakkori et C. Teddlie (dir.), *Handbook of mixed methods in social and behavioral research* (p. 3-50). Thousand Oaks, CA: Sage Publications.

Teddlie, C. et Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Thousand Oaks, CA: Sage Publications.

Teddlie, C. et Tashakkori, A. (2010). Overview of contemporary issues in mixed methods research. Dans A. Tashakkori et C. Teddlie (dir.), *Handbook of mixed methods in social and behavioral research* (p. 1-41). Thousand Oaks, CA: Sage Publications.

Tyson, L. (2014). *Critical theory today*. New York: Routledge.

Van de Ven, A. H. (1992). Suggestions for studying strategy process: a research note. *Strategic Management Journal*, 13(5), 169-188.

Youngs, H. et Piggot-Irvine, E. (2011). The application of a multiphase triangulation approach to mixed methods: The research of an aspiring school principal development program. *Journal of mixed methods research*, 6(3), 184-198.

Résumé / Abstract / Resumen

Les méthodes mixtes (MM) sont de plus en plus populaires. En utilisant les MM, les chercheuses et chercheurs intègrent des méthodes qualitatives (QUAL) et quantitatives (QUAN), par exemple au niveau des questions et devis de recherche, et des collectes et analyses des données. Plusieurs stratégies

d'intégration ont été proposées, mais leur conceptualisation est généralement limitée, fragmentée, ou non testée scientifiquement. Cela constitue un défi pour planifier et effectuer des études utilisant les MM, et pour enseigner les MM. Fondé sur la littérature méthodologique, ce chapitre présente un cadre conceptuel incluant tous les types d'intégration connus, des stratégies pratiques et des combinaisons possibles de stratégies pour intégrer les phases, les résultats et les données QUAL et QUAN. Dans le chapitre suivant, ce cadre est testé avec une revue des études utilisant les MM. Cette revue illustre plusieurs types de combinaisons avec des études en santé mondiale. Ensemble, ces deux chapitres contribuent à faire avancer les connaissances méthodologiques par (a) un appel à une meilleure description des MM dans les publications scientifiques, et (b) une conceptualisation testée proposant trois types d'intégration et neuf stratégies spécifiques, et toutes les combinaisons possibles des stratégies pour intégrer les phases, résultats et données QUAL et QUAN.

***

Mixed methods (MM) are increasingly popular. In performing MM, researchers integrate qualitative (QUAL) and quantitative (QUAN) methods with respect to worldviews, methodologies and designs, research questions, data collection and analysis techniques, and study results. Several integration strategies have been proposed, but their conceptualization is usually design-driven, or fragmented, or not empirically tested. This is challenging for planning and conducting MM studies, and for training graduate students. Based on the methodological literature, this chapter presents a conceptual framework including types of integration, practical strategies, and possible combinations of strategies for integrating QUAL and QUAN phases, results and data. In the next chapter, this framework is tested using a review of 2015 MM studies with a method-detailed description. This supports the framework, illustrates multiple types of combinations, and provides global health examples. These two companion chapters contribute to advance methodological knowledge on MM via (a) a call for better reporting MM studies in scientific articles, and (b) a tested conceptualisation comprising three types of integration and nine specific strategies, which explain current and future possibilities for combining strategies to integrate QUAL and QUAN phases, results, and data.

***

Los métodos mixtos (MM) son cada vez más populares. Al utilizar los métodos mixtos, los investigadores integran métodos cualitativos (QUAL) y cuantitativos (QUAN) con respecto a las cosmovisiones, metodologías y diseños, preguntas de investigación, técnicas de recopilación y análisis de datos y resultados de estudios. Se han propuesto varias estrategias de integración, pero su conceptualización suele estar impulsada por el diseño, o ser fragmentada, o no haber sido probada empíricamente. Esto es un reto para la planificación y realización de estudios que utilizan métodos mixtos y para la formación de estudiantes de postgrado. Basado en la literatura metodológica, este capítulo presenta un marco conceptual que incluye tipos de integración, estrategias prácticas y posibles combinaciones de estrategias para integrar las fases, resultados y datos de QUAL y QUAN. En el siguiente capítulo, este marco se prueba utilizando una revisión de los estudios que utilizan métodos mixtos de 2015 con una descripción detallada del método. Esto apoya el marco, ilustra múltiples tipos de combinaciones y proporciona ejemplos de salud mundial. Estos dos capítulos complementarios contribuyen a avanzar en el conocimiento metodológico de los métodos mixtos a través de (a) un llamado a informar mejor sobre los estudios que utilizan métodos mixtos en artículos científicos, y (b) una conceptualización probada que comprende tres tipos de integración y nueve estrategias específicas, que explican las posibilidades actuales y futuras de combinar estrategias para integrar las fases, los resultados y los datos de QUAL y QUAN.

***

Pierre Pluye est professeur titulaire au Département de médecine de famille de l'Université McGill, chercheur-boursier senior du Fonds de recherche du Québec Santé (FRQS), et directeur de la composante Développement méthodologique à l'Unité SOUTIEN SRAP du Québec. Il est membre associé de l'École des sciences de l'information de l'Université McGill. Il est spécialiste en médecine familiale et en santé publique. Il détient un doctorat en promotion de la santé (Université de Montréal). Il a obtenu des bourses de carrière des Instituts de recherche en santé du Canada (IRSC) et du FRQS. En 2017, il a été élu membre de l'Académie canadienne des sciences de la santé, et a reçu le prix du « Chercheur de l'année » du Collège des médecins de famille du Canada. Il possède une expertise en méthodes mixtes. Ses recherches étudient les effets des informations issues de ressources électroniques sur les professionnels, les patients et le public.

Citation

Pierre Pluye. (2019). L'intégration en méthodes mixtes. Cadre conceptuel pour l'intégration des phases, résultats et données qualitatifs et quantitatifs. In *Évaluation des interventions de santé mondiale. Méthodes avancées*. Sous la direction de Valéry Ridde et Christian Dagenais, pp. 187-212. Québec : Éditions science et bien commun et Marseille : IRD Éditions.

8. La pratique de l'intégration en méthodes mixtes

Les multiples combinaisons des stratégies d'intégration

PIERRE PLUYE, ENRIQUE GARCÍA BENGOECHEA, DAVID LI TANG, VERA GRANIKOV

L'intégration en méthodes mixtes

Définition de la méthode

Les méthodes mixtes intègrent des méthodes qualitatives et quantitatives. Elles permettent de répondre à des questions de recherche et d'évaluation complexes.

Force de la méthode

Nous proposons trois types d'intégration et neuf stratégies pratiques, et toutes les combinaisons possibles des stratégies pour intégrer les phases, résultats et données QUAL et QUAN ; ces propositions sont illustrées avec des exemples issus des recherches en santé mondiale.

Défis de la méthode

Notre proposition et ces exemples peuvent combler un besoin (a) des chercheuses et chercheurs pour mieux planifier, effectuer et publier des études utilisant les méthodes mixtes, et (b) des enseignant-e-s pour former les étudiant-e-s en méthodes mixtes.

Les méthodes mixtes (MM) sont définies comme l'ensemble des combinaisons des méthodes qualitatives (QUAL) et quantitatives (QUAN) en évaluation de programme, en recherche primaire et en revue de littérature. Elles ont émergé dans les années 1950 en santé mondiale (Pluye, 2012). Dans le chapitre précédent, nous avons résumé les connaissances de base en méthodes mixtes et proposé un cadre conceptuel des techniques d'analyse en méthodes mixtes. Dans ce chapitre, nous testons et illustrons cette conceptualisation avec des recherches en santé mondiale.

Contexte

Notre cadre conceptuel inclut trois types d'intégration, neuf stratégies spécifiques et sept combinaisons de stratégies pour intégrer les phases,

résultats et données QUAL et QUAN. Il suggère que toute étude utilisant des MM peut combiner plusieurs stratégies. Le premier type d'intégration est la connexion de phases et correspond à trois stratégies spécifiques : (1.1) la connexion d'une phase QUAL à une phase QUAN; (1.2) la connexion d'une phase QUAN à une phase QUAL; et (1.3) le cas particulier des stratégies précédentes qui est appelé le « suivi d'un fil d'Ariane » entre les phases QUAL et QUAN. Le deuxième type d'intégration est la comparaison de résultats et réfère à trois stratégies spécifiques : (2.1) la comparaison des résultats obtenus de manière séparée; (2.2) la comparaison des résultats obtenus de manière interdépendante; et (2.3) le cas particulier des stratégies précédentes qui est appelé « divergence » des résultats QUAL et QUAN. Le troisième et dernier type d'intégration est l'assimilation des données. Il inclut aussi trois stratégies spécifiques qui sont : (3.1) la transformation de données QUAL en données QUAN (quantification); (3.2) la transformation de données QUAN en données QUAL (qualitativation); et (3.3) la fusion des données QUAL et QUAN. Pour les raisons que nous avons expliquées dans le précédent chapitre, l'ensemble des combinaisons possibles de ces stratégies (1 ; 2 ; 3 ; 1 et 2 ; 1 et 3 ; 2 et 3 ; 1 et 2 et 3) constitue un nouveau monde à explorer en méthodes mixtes (une combinaison étant définie par un ou plusieurs éléments d'un ensemble).

Plan

Pour tester et illustrer cette conceptualisation avec des recherches en santé mondiale récentes, nous avons utilisé eSRAP qui est un système de veille de tendances en recherche (encadré 1) (Tang, Pluye et Bouthillier 2015). eSRAP permet la collaboration des membres de communautés en recherche axée sur le patient (RAP) pour évaluer et partager les résultats et les projets de recherche de manière adaptée aux besoins des usagers. Selon les Instituts de recherche en santé du Canada (IRSC), la recherche axée sur le patient fait référence à un continuum de la recherche qui engage les patients en tant que partenaires, met l'accent sur les priorités des patients, et améliore les services de santé et la santé. Elle est multidisciplinaire et conduite en partenariat avec toutes les parties prenantes. Le mot « patient » représente la population en général, incluant les personnes avec un problème social ou de santé et leurs aidants.

Les auteurs et autrices de ce chapitre ont contribué au développement du prototype d'eSRAP qui a été spécifiquement conçu pour faciliter la veille des tendances en recherche aussitôt que les publications sont indexées dans des bases de données bibliographiques. Au moment où ce chapitre est écrit, 15 chercheuses et chercheurs (enseignant-e-s et anciens étudiant-e-s des cours en méthodes mixtes offerts par le département de médecine familiale de l'Université McGill depuis 2008) utilisent eSRAP pour analyser et comprendre l'environnement et les tendances émergentes sur le plan des stratégies pour intégrer les phases, les résultats et les données QUAL et QUAN en méthodes mixtes.

Encadré 1. eSRAP : veille collaborative des tendances en recherche axée sur le patient

Pourquoi les communautés RAP ont besoin d'eSRAP?

eSRAP est un système de Veille des tendances en recherche (STR) qui permet la collaboration des membres de communautés en recherche axée sur le patient (RAP) pour analyser et comprendre l'environnement et les tendances émergentes dans leurs domaines de recherche de manière adaptée à leurs besoins. eSRAP vise à mettre les communautés RAP au courant de l'avancement des connaissances dans leurs domaines respectifs. Il facilite la veille des développements d'avant-garde aussitôt qu'ils apparaissent sous forme de publications indexées dans les bases de données bibliographiques.

Qu'est-ce qu'un système STR ?

La STR applique les principes de la veille concurrentielle et de la surveillance de l'environnement pour mettre à jour de manière rapide, constante et structurée les connaissances les plus récentes. La STR peut compléter les revues de littérature systématiques, et remplacer les mécanismes d'alerte traditionnels. D'une part, les revues systématiques sont très complètes et produisent des recommandations précieuses, mais sont coûteuses, *post hoc* et ne sont pas constamment à jour. D'autre part, les systèmes d'alerte traditionnels (fils RSS) fournissent une simple liste de titres et résumés non classifiés et souvent non pertinents.

Pour sa part, la STR fournit une sélection des documents pertinents de qualité et des résultats d'analyse, par et pour une communauté d'usagers. La STR peut nécessiter peu de ressources (les ressources nécessaires diminuent lorsque le nombre de membres de la communauté augmente) et conserve les avantages de capitaliser et d'avoir un effet de levier sur un corpus toujours plus grand et grandissant des connaissances et projets scientifiques. De plus, la STR est axée vers l'avenir, car elle focalise sur les plus récents protocoles et publications au fur et à mesure qu'ils sont publiés.

Chaque partie d'eSRAP est adaptable aux besoins d'une communauté RAP

La stratégie de veille est déterminée avec les experts du sujet/domaine (par exemple, les leaders d'une communauté RAP). Une approche structurée pour diffuser et organiser les nouvelles connaissances est validée et opérationnalisée. Les analyses peuvent être automatisées (avec algorithmes), semi-automatisées ou manuelles. Les alertes peuvent être créées pour la dissémination des nouvelles connaissances et des résultats d'analyse aux membres de la communauté. La communauté RAP peut utiliser ce répertoire structuré de connaissances (continuellement mis à jour) pour générer des analyses qui peuvent faciliter des formations, des protocoles et des publications dans les médias traditionnels et sociaux. eSRAP est financé par l'Unité SOUTIEN SRAP du Québec et développé par des expert-e-s académiques et industriels en veille concurrentielle et en surveillance de la science et des technologies.

Référence : Tang D., Pluye P. et Bouthillier F. (2015). eSRAP : Système de surveillance des tendances en recherche permettant la collaboration des membres de communautés en recherche axée sur le patient (RAP) pour évaluer et partager les résultats et les projets de recherche de manière adaptée. Registration of Copyright (# 1126124), Canadian Intellectual Property Office, Industry Canada.

Pour écrire ce chapitre, eSRAP a permis de sélectionner des articles qui décrivent de manière détaillée les stratégies utilisées dans les recherches axées sur le patient en méthodes mixtes dans les sciences de la santé et les sciences sociales. Nos critères d'éligibilité étaient les suivants : étude

empirique utilisant les méthodes mixtes dans des recherches axées sur le patient publiées en 2015. Notre source d'information était la base de données bibliographiques Scopus. Nous avons utilisé la requête suivante : « TITLE(mixed PRE/5 method*) AND SUBJAREA(MEDI OR SOCI) ». Cela permettait d'identifier un nombre de documents (auteurs et autrices, journal, année, titre et résumé) liés aux MM avec une précision élevée (faible rappel) qui soit gérable par deux réviseurs (Pierre Pluye et Enrique García Bengoechea). Nous avons sélectionné les documents et les articles pertinents en utilisant un manuel de codage incluant la définition des méthodes mixtes et de la recherche axée sur le patient.

Nous avons inclu uniquement les études utilisant les méthodes mixtes (MM) et fournissant une description détaillée de cette utilisation. Nous nous sommes inspirés des six recommandations GRAMMS (*Good Reporting of A Mixed Methods Study*) pour définir ce critère de sélection (O'Cathain, 2010). Ces recommandations sont de décrire au moins :

1. l'objectif et le devis de MM;
2. l'échantillonnage, la collecte et l'analyse des données QUAL et QUAN;
3. l'intégration des phases, résultats et données QUAL et QUAN;
4. une justification pour utiliser les MM;
5. une description des gains associés avec l'intégration des méthodes;
6. une description des limites de cette intégration.

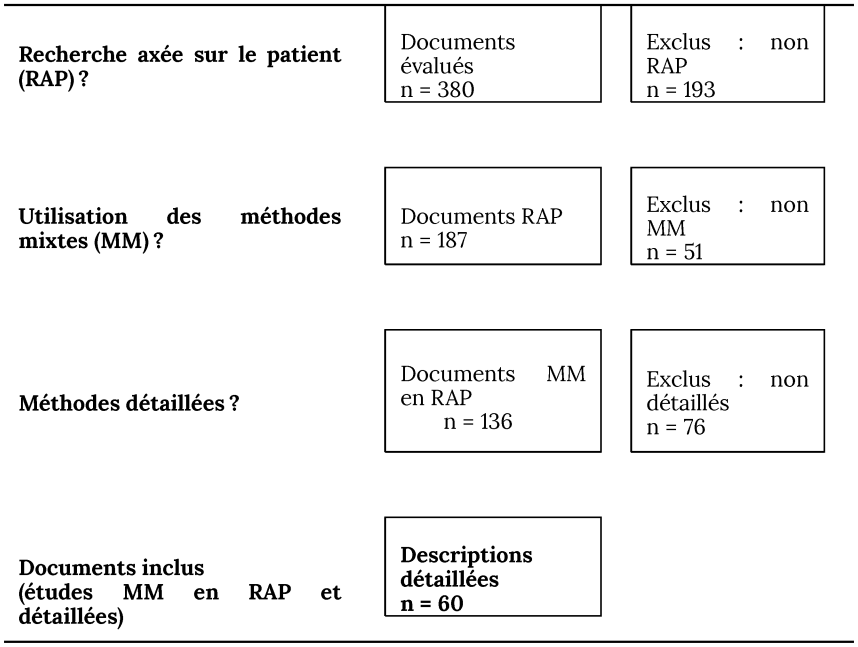
Pour satisfaire ce critère de sélection (description détaillée), nous avons jugé que, minimalement, les trois composantes principales des MM (méthodes QUAL et QUAN et intégration des méthodes QUAL et QUAN) devaient satisfaire les trois recommandations GRAMMS (recommandations 1-2-3). Nous avons utilisé les trois dernières recommandations GRAMMS (recommandations 4-5-6) pour décrire les études inclues, mais non comme critères d'exclusion.

Application

Parmi les 380 documents intitulés « méthodes mixtes » évalués au moment de soumettre ce chapitre, 187 (49,2 %) concernaient la RAP (incluant 35 en santé mondiale) (Figure 1). Parmi ces derniers, 136 (35,8 %) rapportaient

des recherches empiriques primaires qui satisfaisaient la définition des MM (incluant 29 en santé mondiale). Uniquement 60 documents (15,8 %), incluant 11 études en santé mondiale, fournissaient une description détaillée des trois composantes principales des MM et satisfaisaient ainsi les trois premières recommandations GRAMMS (recommandations 1-2-3). Les autres documents ne décrivaient pas la méthodologie et les méthodes, ou le faisaient en une phrase ou un paragraphe, ou décrivaient uniquement les composantes QUAL et QUAN.

Figure 1. Diagramme de flux eSRAP-MM



Parmi les 51 documents qui ne satisfaisaient pas notre définition des MM (incluant six en santé mondiale), 20 rapportaient une collecte ou analyse d'informations qualitatives qui n'étaient pas des « données de recherche », car elles ne référaient pas à une méthodologie et une méthode QUAL (incluant deux en santé mondiale) et 19 documents rapportaient une collecte ou analyse d'informations quantitatives qui n'étaient pas des « données de recherche », car elles ne référaient pas à un devis et une méthode QUAN (incluant quatre en santé mondiale); un document rapportait une juxtaposition de méthodes QUAL et QUAN sans aucune intégration (comme deux études séparées), un autre rapportait une combinaison de méthodes

QUAN, un autre ne rapportait aucune méthodologie ou méthode (description de programme) et neuf étaient non-éligibles pour diverse raisons (revue de littérature, texte complet confus ou indisponible via la bibliothèque de McGill).

Les résultats suivants portent uniquement sur l'analyse des 60 publications qui rapportaient des études RAP satisfaisant la définition des MM et qui fournissaient une description détaillée des trois composantes principales des MM satisfaisant les trois premières recommandations GRAMMS (recommandations 1-2-3). Dans cet échantillon, 38 études (63,3 %) suivaient un devis convergent, 17 (28,3 %) un devis séquentiel, quatre (6,7 %) une devis multiphases et une étude (1,7 %) suivait un devis multiniveaux. Parmi les 60 documents retenus, uniquement six (10,0 %) satisfaisait les trois dernières recommandations GRAMMS (recommandations 4-5-6), 18 (30,0 %) en satisfaisaient deux, 25 (41,7 %) une seule et 11 (18,3 %) n'en satisfaisaient aucune. Toutes les stratégies spécifiques et combinaisons de stratégies proposées sont présentées dans le tableau 1 et illustrées dans la suite du chapitre avec des exemples en santé mondiale.

Tableau 1. eSRAP-MM : Distribution des publications MM en RAP selon les types de devis et les combinaisons de stratégies (N=60)

3 TYPES D'INTÉGRATION 9 TYPES DE STRATÉGIES	COMBINAISONS DE STRATÉGIES SPÉCIFIQUES AUX MM (SELON LES TYPES DE DEVIS) Symbole de la combinaison (n = nombre d'études illustrant chaque combinaison)			
	Séquentiel	**Convergent**	**Multiphases**	**Multiniveaux**
1. CONNEXION DES PHASES				
1.1 Phase QUAL à phase QUAN	1.1 (n=1)	na	1.1 (n=1)	
1.2 Phase QUAN à phase QUAL	1.2 (n=3)	na		
1.3. Cas particulier : Suivre le fil (follow a thread)		na	1.1 et 1.2 et 1.3 et 2.2 (n=1)	
2. COMPARAISON DES RÉSULTATS				
2.1 Résultats QUAL et QUAN obtenus de manière séparée	s.o.	2.1 (n=12)	na	
2.2 Résultats QUAL et QUAN obtenus de manière interdépendante	1.2 et 2.2 (n=7)	2.2 (n=18)	1.1 et 2.2 (n=2)	
2.3. Cas particulier : Divergence	1.1 et 2.3 (n=2) 1.2 et 2.2 et 2.3 (n=1)	2.1 et 2.3 (n=2)		
3. ASSIMILATION DES DONNÉES				
3.1 Données QUAL en données QUAN	1.1 et 3.1 (n=1)	2.2 et 3.1 (n=2) 3.1 (n=1)		3.1 (n=1)
3.2 Données QUAN en données QUAL		3.2 (n=1)		

3.3 Fusion des données QUAL et QUAN	1.1 et 2.2 et 3.3 (n=1) 1.2 et 3.3 (n=1)	2.2 et 3.3 (n=2)		
4. STRATÉGIES ÉMERGENTES				
(...)				
Nombre de combinaisons observées	**4 (n=17)**	**3 (n=38)**	**2 (n=4)**	**1 (n=1)**

**Les cas particuliers ne sont pas inclus dans le dénombrement des combinaisons plausibles.*

Note : Le logogramme « et » (signifiant « et ») permet de représenter les combinaisons avec plusieurs stratégies.

Intégration 1. Connexion de phases

Stratégie 1.1. Connexion d'une phase QUAL à une phase QUAN

Fenenga et ses collègues ont utilisé 42 entretiens avec des familles et 7 entretiens avec des informateurs clés (phase-1 : données QUAL), et un questionnaire structuré complété par un échantillon aléatoire représentatif des familles de 64 régions (phase-2 : données QUAN ; n = 1903), pour analyser l'influence des relations sociales sur les perceptions, le comportement et les processus de prise de décisions des familles à l'égard de leur intention de joindre le plan national d'assurance maladie au Ghana (Fenenga *et al.* 2015). Les résultats des analyses des données QUAL (théorisation ancrée) ont informé (connexion de phases) le développement du questionnaire pour la collecte des données QUAN dans la deuxième phase de l'étude (enquête transversale). Ce faisant, les chercheuses et chercheurs ont développé des questions ancrées dans la réalité des familles (validité écologique de contenu). Les résultats des analyses des données QUAN (régression logistique) ont ensuite permis de généraliser les résultats QUAL de manière statistique, par exemple l'importance et la probabilité des associations entre les variables étudiées. En plus de la connexion de phases, les résultats QUAL et QUAN ont été comparés (voir plus loin : « cas particulier des stratégies 2.1

et 2.2 »), ce qui illustre une combinaison de stratégies (connexion de phases et comparaison de résultats).

Stratégie 1.2. Connexion d'une phase QUAN à une phase QUAL

De Allegri et ses collègues ont étudié les facteurs associés à la prise de décision des hommes habitant les régions rurales du Burkina Faso concernant le passage du test de dépistage du virus du SIDA (De Allegri *et al.* 2015). Un échantillon représentatif des hommes adultes a rempli un questionnaire structuré validé (n=1058). Les résultats des analyses des données QUAN de la phase-1 de l'étude (statistiques descriptives et analyses multivariées) ont informé la stratégie d'échantillonnage raisonné pour planifier la phase-2 QUAL de l'étude. Pour cette phase, des entretiens semi-structurés ont été effectués avec 38 hommes (19 testés et 19 non testés) recrutés en fonction de leur âge et de leur région. Les résultats des analyses descriptives (inductives) des données QUAL ont permis de mieux comprendre et interpréter les résultats de la phase-1 QUAN.

Cas particulier des stratégies 1.1 et 1.2 : Suivi d'un fil d'Ariane entre phases QUAL et QUAN

L'utilisation de cette stratégie est illustrée dans une étude effectuée par Boot et ses collègues (Boot *et al.* 2016). L'objectif des chercheuses et chercheurs était de mieux comprendre les différences et les similarités entre les facteurs associés à la participation au travail des travailleurs hollandais âgés de 58 à 65 ans et souffrant d'au moins un des trois problèmes de santé chroniques suivantes : dépression, maladie cardiovasculaire et ostéoarthrite. Les données QUAN provenaient d'une étude de cohorte populationnelle (*Longitudinal Aging Study Amsterdam*; n=1002). Les chercheuses et chercheurs ont d'abord procédé à des analyses statistiques (tests de t et du chi-carré) pour tester leurs hypothèses initiales et comparer les résultats entre quatre groupes de travailleurs (dépression vs maladie cardiovasculaire vs ostéoarthrite vs combinaison d'au moins deux de ces problèmes de santé). Cela a permis de développer un guide d'entrevue pour la composante QUAL de l'étude qui a exploré les possibilités de facteurs émergents (n=14). Les résultats de l'analyse des données QUAL (thématique) ont permis aux chercheuses et chercheurs de découvrir et proposer des nouvelles variables

(prédicteurs émergents – non prévus initialement). Ces variables (fil d'Ariane) ont été utilisées pour générer de nouvelles hypothèses et effectuer des analyses statistiques additionnelles.

Intégration 2. Comparaison de résultats

Stratégie 2.1. Comparaison des résultats obtenus de manière séparée

Cette stratégie a été utilisée par Pfadenhauer et Rehfuess dans une étude qui visait à évaluer et améliorer les comportements de la population concernant la qualité de l'eau, les installations sanitaires et le lavage des mains dans les écoles d'une région des Philippines (Pfadenhauer et Rehfuess 2015). Cette étude a évalué l'impact de ces comportements sur la présence de diarrhée infantile en tenant compte de l'influence des facteurs socioculturels et environnementaux. À cette fin, les chercheuses ont analysé (régression logistique) les données QUAN provenant d'une enquête transversale nationale (échantillon représentatif de la population; n = 12 469 foyers). De manière concomitante, elles ont organisé six groupes de discussion avec des femmes habitant des zones urbaines et rurales de la région. Les données QUAL ont été analysées de manière inductive (analyse thématique) pour produire un cadre conceptuel des facteurs associés aux comportements concernés. Les données QUAN et QUAL ont été analysées et présentées séparément. Dans la section « Discussion » de leur article, les chercheuses ont comparé les résultats QUAN et QUAL (matrice des résultats) et interprété cette comparaison sous forme de méta-thèmes communs aux résultats QUAL et QUAN en identifiant les similarités (complètes/partielles), les différences (présence/absence) et les divergences entre les résultats QUAN et QUAL (voir « cas particulier des stratégies 2.1 et 2.2 » ci-dessous).

Stratégie 2.2. Comparaison des résultats obtenus de manière interdépendante

Oguntunde et ses collègues ont publié un exemple de cette stratégie dans le contexte d'une étude visant à examiner les facteurs (facilitateurs et barrières) associés à l'usage de sulfate de magnésium pour soigner les femmes en pré-éclampsie (hypertension sévère liée à la grossesse) dans

des centres de santé du nord du Nigeria (Oguntunde *et al.* 2015). Dans un échantillon aléatoire de 80 centres, les chercheuses et chercheurs ont effectué une enquête transversale QUAN auprès de 80 cliniciens (un par centre) sur la prise en charge des femmes en pré-éclampsie (connaissances et protocoles de soins, et ressources humaines et matérielles disponibles). De plus, ils ont effectué une recherche QUAL descriptive et interviewé un échantillon raisonné (par exemple, centre rural/urbain) de 30 gestionnaires (un par centre). Les données QUAN ont été analysées avec des statistiques descriptives alors que les données QUAL ont été analysées de manière thématique (les thèmes étant les facteurs étudiés). Dans la section des résultats, les chercheuses et chercheurs ont comparé les résultats QUAN et QUAL de manière systématique pour chaque facteur (groupe thématique). Pour la plupart des facteurs, les résultats sont présentés de manière à ce que les résultats QUAL expliquent et complètent les résultats QUAN. Pour cinq autres facteurs, les résultats présentés sont uniquement QUAN (n=4) ou QUAL (n=1). Selon les chercheuses et chercheurs, cette comparaison leur a permis de mieux comprendre les facteurs étudiés et de proposer des interventions adaptées.

Cas particulier des stratégies 2.1 et 2.2 : Divergence des résultats QUAL et QUAN

Pfadenhauer et Rehfuess (voir « stratégie 2.1 » ci-dessus) disent avoir créé une « matrice de convergence » (non présentée dans leur article) dans laquelle elles ont porté une attention spécifique aux divergences entre les résultats QUAN et QUAL. Dans l'article, ces divergences semblent réconciliées sous forme d'interprétation des résultats QUAN par des résultats QUAL (nuances). De façon similaire, Fenenga et ses collègues (voir plus haut « stratégie 1.1 ») ont comparé leurs résultats QUAL et QUAN dans un tableau et présenté de façon détaillée la cohérence entre ces résultats et une divergence. Contrairement à ce que suggéraient les résultats QUAL, l'engagement des participant-e-s dans des groupes d'action sociale n'était pas statistiquement associé avec l'intention de joindre le plan national d'assurance maladie au Ghana. Les chercheuses et chercheurs ont proposé de réconcilier cette divergence en suggérant que l'action sociale/communautaire était rarement centrée sur la santé dans ce contexte.

Intégration 3. Assimilation des données

Stratégie 3.1. Transformation de données QUAL en données QUAN (quantification)

Valentijn et ses collègues ont utilisé cette stratégie dans le cadre d'une étude portant sur les processus de collaboration et l'efficacité perçue des projets de soins intégrés en Hollande (Valentijn *et al.* 2015). Dans 42 projets, ces deux aspects ont été évalués à trois niveaux : systémique, organisationnel et professionnel. Au niveau du système, les évaluateurs et les évaluatrices ont interviewé les coordinateurs et les parties prenantes des projets (données QUAL). Aux niveaux organisationnel et professionnel, les mesures étaient fondées sur des questionnaires structurés complétés respectivement par les membres des comités de pilotage des projets et par les professionnel-le-s impliqués (données QUAN). Les chercheuses et chercheurs ont utilisé une procédure valide et fiable (analyse thématique suivie d'une analyse quantitative de contenu) pour transformer les données QUAL en données QUAN. Ces dernières (données QUAL quantifiées) ont ensuite été utilisées avec les autres données QUAN pour effectuer des analyses statistiques (regroupement hiérarchique) et comparer les processus de collaboration et l'efficacité perçue des projets de soins intégrés selon les trois niveaux prévus (systémique, organisationnel et professionnel).

Stratégie 3.2. Transformation de données QUAN en données QUAL (qualitativation)

Reichwein et ses collègues ont entrepris une étude-pilote dont le but était de planifier un programme de promotion de la santé en planification familiale (Reichwein *et al.* 2015). Spécifiquement, les chercheuses et chercheurs visaient l'identification des usagers et usagères ayant le plus besoin de méthodes contraceptives en Ouganda et au Vietnam afin d'orienter le programme à leur bénéfice. Les chercheuses et chercheurs ont analysé de manière statistique (regroupement hiérarchique) les données QUAN collectées lors d'enquêtes nationales transversales et identifié trois profils d'usagères et d'usagers potentiels : des femmes ougandaises voulant rallonger les périodes entre leurs grossesses, des hommes ougandais désirant limiter le nombre de leurs enfants et des jeunes femmes

vietnamiennes sexuellement peu actives. Pour chacun de ces groupes, des données QUAL ont été collectées (entrevues individuelles et groupes de discussion) et analysées de manière thématique. Les chercheuses et chercheurs ont ensuite assimilé les données QUAN transformées en profils narratifs et les données QUAL (types d'histoires de cas) pour créer deux archétypes biographiques (deux idéaux types) sur lesquels centrer la planification du programme : « *Kibuuka* », un fermier de 52 ans semi-analphabète habitant une région rurale en Ouganda et « *Anh* », une étudiante de 20 ans habitant Hanoi, la capitale du Vietnam. Les biographies idéales typiques ainsi développées personnifient les caractéristiques et croyances propres aux deux groupes cibles du programme. La qualitativation a permis d'assimiler des profils statistiques (qualitatifs) et des données QUAL sous forme d'archétypes (idéaux-types).

Stratégie 3.3. Fusion des données QUAL et QUAN

Johansson et ses collègues ont examiné l'effet des tests de dépistage sur l'utilisation des médicaments chez des enfants fiévreux dans des pays africains où la malaria est endémique (Johansson *et al.* 2015). Dans 12 pays, ils ont collecté des données QUAN issues d'enquêtes nationales sur les services de santé et la santé des enfants (n=16 323) et analysé ces données pour chaque pays (régression logistique). Concernant six de ces pays, les chercheuses et chercheurs ont utilisé un devis d'étude de cas multiples (chaque pays étant un cas) et collecté des données QUAL via des documents sur les programmes de lutte contre la malaria et des entrevues avec sept expert-e-s (au total) des programmes et pays concernés. Pour chaque pays (cas), les résultats QUAN et QUAL ont été regroupés ensemble dans une matrice (les résultats QUAL ayant permis de développer quatre explications -thèmes- principales plausibles des résultats QUAN). Une méta-matrice incluant toutes les données de tous les cas a été développée et constituait en soi une nouvelle base de données qui a permis d'évaluer les effets étudiés dans chaque pays (chaque étude de cas est présentée dans un fichier supplémentaire en annexe) en plus de révéler des similarités et des différences entre les pays (comparaison des cas).

Combinaisons de stratégies et stratégies émergentes

Le tableau 1 présente les multiples combinaisons de stratégies observées dans cet échantillon de 60 études MM en santé mondiale. Toutes les combinaisons possibles sont représentées (1 ; 2 ; 3 ; 1 et 2 ; 1 et 3 ; 2 et 3 ; 1 et 2 et 3). Environ la moitié des combinaisons sont représentées pour les devis séquentiels et convergents. Étant donné la rareté des devis multiphases et multiniveaux, peu de combinaisons ont été observées pour ces devis. Aucune stratégie ou combinaison émergente (imprévue) n'a été observée.

Analyse réflexive et leçons à tirer

Ces résultats illustrent le cadre conceptuel proposé : trois types d'intégration, neuf stratégies spécifiques et sept combinaisons de stratégies pour intégrer des phases, résultats et données QUAL et QUAN en méthodes mixtes. Bien que notre conceptualisation et le système de veille eSRAP permettent d'identifier des stratégies émergentes, nous n'en avons détecté aucune jusqu'à maintenant. La majorité des exemples présentés dans ce chapitre proviennent des études MM en santé mondiale et démontrent le dynamisme des chercheuses et chercheurs dans ce domaine. Ce chapitre contribue aux connaissances en MM en proposant une conceptualisation opérationnelle (pratique) des types d'intégration et des stratégies spécifiques pour intégrer les phases, résultats et données QUAL et QUAN. Concernant les innovations, il montre la richesse des combinaisons de stratégies en MM pour chaque devis (tableau 1).

Nous avons surveillé uniquement un échantillon des études utilisant les MM en RAP publiées en 2015, sans chercher à fournir une description exhaustive des recherches en santé mondiale utilisant les MM. Quoique limités sur le plan de l'exhaustivité, ces résultats montrent que eSRAP offre l'avantage à une communauté grandissante de chercheuses et chercheurs en MM de se tenir au courant des nouvelles tendances de manière prospective en partageant le fardeau de la sélection des études utilisant les MM et fournissant une description détaillée de cette utilisation. Spécifiquement, eSRAP est adaptable aux besoins individuels des évaluateurs et évaluatrices, des chercheuses et chercheurs et des enseignant-e-s, et leur permet de se tenir à jour sur un sujet de leur choix sans s'épuiser. Par exemple, chaque

membre de la communauté des eSRAPeurs en MM peut bénéficier à chaque instant des résultats du travail prospectif de tous les membres. Pour les enseignant-e-s en MM, eSRAP permet en quelques clics avant chaque cours de retenir des études récentes comme matériel pédagogique pour les étudiant-e-s via les fonctions « filtres ». Par exemple, eSRAP permet de trouver rapidement des articles récents illustrant chaque stratégie de manière détaillée.

Ce chapitre a montré qu'une proportion substantielle des études intitulées MM correspondent en fait à des méthodes QUAL (10 %) ou QUAN (10 %). En 2016, tous les chercheuses et chercheurs s'entendent pour dire que les MM combinent des méthodes QUAL et QUAN (Plano Clark et Ivankova 2016) et nous avons inclus des études où les chercheuses et chercheurs « collectent et analysent de manière convaincante et rigoureuse des données QUAN et QUAL » (Creswell et Plano Clark, 2011, p. 5). Cette rigueur dépend de l'expertise des chercheuses et chercheurs, notamment de leur choix parmi (a) les instruments de mesure (validité et fidélité) et de devis (essai contrôlé randomisé vs étude non randomisée vs étude descriptive) pour les données QUAN et parmi (b) les interactions avec les participant-e-s (crédibilité et contextualisation) et de devis (étude descriptive ou interprétative, étude de cas, théorisation ancrée, biographie ou ethnographie entre autres) pour les données QUAL (Creswell et Plano Clark 2011). En appliquant ce critère de la définition des MM, nous avons exclu une étude RAP sur cinq intitulée MM publiée en 2015 (39/187 = 20,9 % au total ; 6/35 = 17,1 % en santé mondiale).

Ces études exclues peuvent être interprétées comme un signe de la reconnaissance des MM dans la communauté scientifique et une illustration des extrêmes du continuum entre méthodes QUAL et QUAN. En premier lieu, la reconnaissance nouvellement acquise des MM suggère que le faible nombre de réviseurs, réviseuses ou d'éditeurs, d'éditrices experts en MM permettait encore en 2015 de publier des études intitulées « méthodes mixtes » qui ne satisfont pas les critères de la définition la plus fréquemment citée des MM et acceptée par la communauté des chercheuses et chercheurs en MM, notamment les membres de MMIRA (*Mixed Methods International Research Association*) (Creswell et Plano Clark 2011; Johnson, Onewuegbuzie, et Turner 2007). Autrement dit, la proportion substantielle de ces études suggère que les MM exercent un attrait sur les chercheuses et chercheurs

en recherche axée sur le patient et les éditeurs et éditrices (biais positif de publication).

En deuxième lieu, le continuum entre méthodes QUAL et QUAN est centré sur les MM (Figure 2) (Johnson *et al.* 2007; Niglas 2010). Nos résultats suggèrent qu'entre méthodes QUAL et MM se trouvent des recherches QUAL intitulées MM où les participant-e-s sont décrits avec des chiffres (entre autres informations). Nous avons exclu 19 études pour cette raison (présence d'information quantitative, mais absence de données QUAN satisfaisant les critères minimum de rigueur ci-dessus). De l'autre côté du continuum, entre MM et méthodes QUAN se trouvent des enquêtes QUAL intitulées MM avec quelques mots dans une boite de commentaires optionnelle en fin de questionnaire anonyme auto-administré. Nous avons exclu 20 études pour cette raison (présence d'information qualitative, mais absence de données QUAL satisfaisant les critères minimum de rigueur).

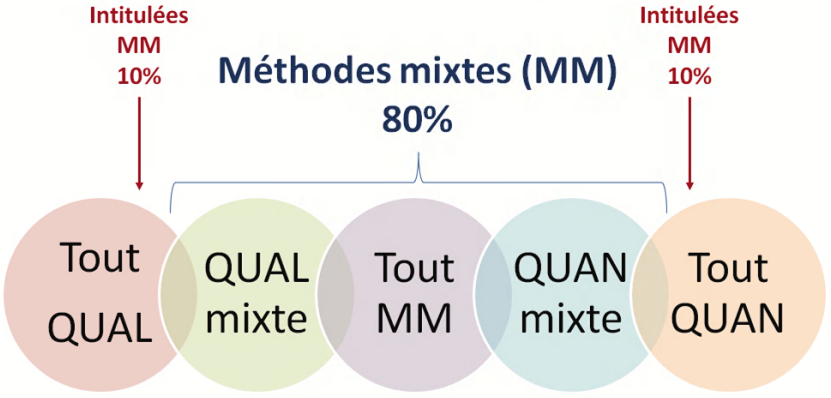


Figure 2. Recherche axée sur le patient en 2015 : études qualitatives (QUAL) ou quantitatives (QUAN) intitulées méthodes mixtes (MM) sans satisfaire la définition des MM (Source : Johnson et al. 2007)

La proportion des études intitulées MM qui ne satisfont pas les critères de la définition usuelle des MM devrait stagner ou diminuer si les « instructions aux auteurs et autrices » des journaux (et les grilles d'évaluation des réviseurs et réviseuses) incluent des recommandations pour rapporter les MM telles que GRAMMS (*Good Reporting of A Mixed Methods*

Study) (O'Cathain, Murphy et Nicholl 2008). Nous rappelons ici que les six recommandations GRAMMS sont de décrire au moins :

1. l'objectif et le devis MM;
2. l'échantillonnage, la collecte et l'analyse des données QUAL et QUAN;
3. l'intégration des phases, résultats et données QUAL et QUAN;
4. une justification pour utiliser les MM;
5. une description des gains associés avec l'intégration des méthodes;
6. une description des limites de cette intégration.

Finalement, huit ans après la publication des six recommandations GRAMMS (pour bien rapporter les études utilisant des MM), nos résultats sont décevants. Plus de la moitié des études MM en RAP (76/136 = 55,9 % dans tout l'échantillon ; 18/29 = 62,1 % en santé mondiale) ne satisfont pas les trois premières recommandations GRAMMS (recommandations 1-2-3), par exemple en raison de l'absence de description des aspects QUAL, QUAN et MM. Trop rares sont les documents qui appliquent toutes les six recommandations GRAMMS (recommandations 1-2-3-4-5-6) : six (6/136 = 4,4 %) dans tout l'échantillon, dont une (1/29 = 3,4 %) en santé mondiale (Akelo *et al.*,2015). Cela suggère de faire appel aux responsabilités conjointes des auteurs et autrices, des réviseurs et réviseuses, des éditeurs et éditrices et des organisations internationales comme l'Association internationale pour les recherches utilisant des méthodes mixtes (MMIRA) afin de promouvoir l'application de ces recommandations en tenant compte des spécificités des disciplines et des revues. Comme tout travail scientifique, les MM doivent être explicites, transparentes et reproductibles. Spécifiquement, mieux décrire les MM sera utile pour aider les étudiant-e-s à mieux comprendre les méthodes et stratégies mobilisées pour intégrer des phases, résultats et données QUAL et QUAN.

Conclusion

Ce chapitre rappelle que les méthodes mixtes requièrent habituellement la collaboration d'expert-e-s en méthodes QUAL et en méthodes QUAN (par exemple une personne experte en QUAL et une autre en QUAN pour les doctorant-e-s) qui doivent réconcilier leurs interprétations des sciences lorsque celles-ci diffèrent et se porter garant-e-s de la rigueur

méthodologique des aspects QUAL et QUAN des MM. De plus, nous mentionnons six recommandations minimales pour bien rapporter une étude utilisant les MM et nous proposons trois types d'intégration, neuf stratégies spécifiques aux MM et tout un nouveau monde à explorer avec sept combinaisons de stratégies pour intégrer des phases, des résultats et des données QUAL et QUAN. Cette conceptualisation contribue aux connaissances sur les MM et se veut pratique. Elle peut guider les étudiant-e-s et les chercheuses et chercheurs pour planifier, conduire et évaluer des études et des évaluations de programmes utilisant les MM, notamment en santé mondiale.

Références clés

Méthodes mixtes francophonie :
http://methodesmixtesfrancophonie.pbworks.com

Cette nouvelle association internationale à but non lucratif vise à promouvoir les méthodes mixtes en français. En 2018, elle regroupe déjà 121 membres de 8 pays différents (Allemagne, Belgique, Canada, France, Rwanda, Sénégal et Suisse). Elle offre un soutien méthodologique, une bibliographie ouverte contenant les publications des membres en méthodes mixtes, et une liste d'experts, de ressources et de formations. Notamment, elle offre un accès gratuit au premier ouvrage collectif méthodologique de référence sur les méthodes mixtes en français (14 chapitres, 266 pages) : Bujold, M., Hong, Q. N., Ridde, V., Bourque, C. J., Dogba, M. J., Vedel, I. et Pluye, P. (2018). *Oser les défis des méthodes mixtes en sciences sociales et sciences de la santé*. Montréal : 117^{e} Cahiers scientifiques de l'ACFAS.

Onwuegbuzie, A. J. et Combs, J. P. (2010). Emergent data analysis techniques in mixed methods research: A synthesis. Dans A. Tashakkori et C. Teddlie (dir.), *Handbook of mixed methods in social and behavioral research* (p. 397-430). Thousand Oaks, CA: Sage Publications.

Il s'agit d'un portrait général des analyses en méthodes mixtes selon 13 dimensions incluant par exemple l'orientation des analyses (cas, variables, processus) et les liens entre les analyses et les autres aspects méthodologiques (philosophie, objectifs et formes de données). Il permet aux évaluateurs et évaluatrices, aux étudiant-e-s et aux chercheuses et chercheurs utilisant les

méthodes mixtes de situer dans une cadre plus général les stratégies d'analyses proposées et illustrées dans ce chapitre sur la santé mondiale.

Système de veille collaborative des tendances en recherche (eSRAP) : www.esrap.ca

Ce site Internet est produit par l'Unité de soutien SRAP du Québec qui soutient la recherche axée sur le patient. eSRAP est une plateforme de veille collaborative conçue pour aider les communautés intéressées par un sujet (incluant des chercheuses et chercheurs, cliniciens, évaluateurs et évaluatrices, gestionnaires, et patients/citoyens partenaires) à se tenir au courant des dernières publications scientifiques sur ce sujet. eSRAP facilite la sélection des publications pertinentes et l'accès aux publications sélectionnées. eSRAP vise à optimiser le partage des efforts de veille et celui des connaissances, et ultimement à démocratiser les veilles, par exemple en méthodes mixtes dans ce chapitre. Une version libre-service sera disponible en 2019, et pourra soutenir toute communauté intéressée par un sujet en santé mondiale.

Références

Akelo, V., McLellan-Lemal, E., Toledo, L., Girde, S., Borkowf, C. B., Ward, L., ... Thomas, T. K. (2015). Determinants and experiences of repeat pregnancy among HIV-positive Kenyan women: a mixed-methods analysis. *PloS one*, 10(6), e0131163.

Boot, C. R., Kruif, A. T. C., Shaw, W. S., Beek, A. J., Deeg, D. J. et Abma, T. (2016). Factors important for work participation among older workers with depression, cardiovascular disease, and osteoarthritis: a mixed method study. *Journal of occupational rehabilitation*, 26(2), 160-172.

Creswell, J. W. et Clark, V. L. P. (2011). *Designing and conducting mixed methods research* (2e éd.). Thousand Oaks, CA: Sage Publications.

De Allegri, M., Agier, I., Tiendrebeogo, J., Louis, V. R., Yé, M., Mueller, O. et Sarker, M. (2015). Factors affecting the uptake of HIV testing among men: a mixed-methods study in rural Burkina Faso. *PloS one*, 10(7), e0130216.

Fenenga, C. J., Nketiah-Amponsah, E., Ogink, A., Arhinful, D. K., Poortinga, W. et Hutter, I. (2015). Social capital and active membership in the Ghana National Health Insurance Scheme-a mixed method study. *International journal for equity in health*, 14(1), 1.

Johansson, E. W., Gething, P. W., Hildenwall, H., Mappin, B., Petzold, M., Peterson, S. S. et Selling, K. E. (2015). Effect of diagnostic testing on medicines used by febrile children less than five years in 12 malaria-endemic African countries: a mixed-methods study. *Malaria journal*, 14(1), 1.

Johnson, R. B., Onewuegbuzie, A. J. et Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2), 112-133.

Niglas, K. (2010). The multidimensional model of research methodology: an integrated set of continua. Dans A. Tashakkori et C. Teddlie (dir.), *Handbook of mixed methods in social and behavioral research* (2e éd., p. 215-236). Thousand Oaks, CA: Sage Publications.

O'Cathain, A. (2010). Assessing the quality of mixed methods research: towards a comprehensive framework. Dans A. Tashakkori et C. Teddlie (dir.), *Handbook of mixed methods in social and behavioral research* (2e éd., p. 531-555). Thousand Oaks, CA: Sage Publications.

O'Cathain, A., Murphy, E. et Nicholl, J. (2008). The quality of mixed methods studies in health services research. *Journal of Health Services Research and Policy*, 13(2), 92-98.

Oguntunde, O., Charyeva, Z., Cannon, M., Sambisa, W., Orobaton, N., Kabo, I. A., ... et Sadauki, H. (2015). Factors influencing the use of magnesium sulphate in pre-eclampsia/eclampsia management in health facilities in Northern Nigeria: a mixed methods study. *BMC Pregnancy et Childbirth*, 15(1), 1.

Pfadenhauer, L. M. et Rehfuess, E. (2015). Towards effective and socio-culturally appropriate sanitation and hygiene interventions in the Philippines: a mixed method approach. *International journal of environmental research and public health*, 12(2), 1902-1927.

Clark, V. L. P. et Ivankova, N. V. (2016). *Mixed methods research: a guide to the field* (vol.). Thousand Oaks, CA: Sage Publications.

Pluye, P. (2012). Méthodes mixtes en santé mondiale. Dans J. C. Suárez-Herrera et M. J. Blain (dir.), *Regards croisés sur la socio-anthropologie réflexive et la recherche en santé mondiale* (p. 129-135). Montréal : 114e Cahiers scientifiques de l'ACFAS.

Reichwein, B., Wolmarans, L., Nantayi, L., Nassali, F., Kakinda, A., Musumba,

D., … et Baatsen, P. (2015). SegWeigh: a mixed-method approach to segmenting potential contraceptive user groups and meeting Family Planning 2020 goals. *International Journal of Gynecology et Obstetrics*, 130, E8-E14.

Tang, D., Pluye, P. et Bouthillier, F. (2015). *eSRAP : système de surveillance de tendances en recherche permettant la collaboration des membres de communautés en recherche axée sur le patient (RAP) pour évaluer et partager les résultats et les projets de recherche de manière adaptée.* Dans M. Bujold, Q. N. Hong, V. Ridde, C. J. Bourque, M. J. Dogba, I. Vedel et P. Pluye (dir.), *Oser les défis des méthodes mixtes en sciences sociales et sciences de la santé* (p. 40). Montréal : 117e Cahiers scientifiques de l'ACFAS.

Valentijn, P. P., Ruwaard, D., Vrijhoef, H. J., de Bont, A., Arends, R. Y. et Bruijnzeels, M. A. (2015). Collaboration processes and perceived effectiveness of integrated care projects in primary care: a longitudinal mixed-methods study. *BMC health services research*, 15(1), 1.

Remerciements

Pierre Pluye détient une bourse de chercheur boursier senior des Fonds de recherche du Québec en santé (FRQS). Le développement du système eSRAP est soutenu par l'Unité Soutien SRAP du Québec (composante Développements méthodologiques).

Les auteurs et autrices remercient les membres fondateurs de la communauté eSRAP-MM qui ont testé le prototype d'eSRAP, notamment tous les nouveaux et anciens étudiant-e-s et enseignant-e-s des cours FMED-672 « Applied Mixed Methods in Health Research », FMED-600 « Mixed Studies Reviews » et FMED 608 « Advanced Mixed Methods » qui sont offerts depuis 2008 par le Département de médecine familiale de l'Université McGill.

Résumé / Abstract / Resumen

Les méthodes mixtes (MM) sont de plus en plus populaires. En utilisant les MM, les chercheuses et chercheurs intègrent des méthodes qualitatives (QUAL) et quantitatives (QUAN), par exemple au niveau des questions et devis

de recherche, et des collectes et analyses des données. Plusieurs stratégies d'intégration ont été proposées, mais leur conceptualisation est généralement limitée, fragmentée, ou non testée scientifiquement. Cela constitue un défi pour planifier et effectuer des études utilisant les MM, et pour enseigner les MM. Fondé sur la littérature méthodologique, le chapitre précédent a présenté un cadre conceptuel incluant tous les types d'intégration connus, des stratégies pratiques et des combinaisons possibles de stratégies pour intégrer les phases, les résultats et les données QUAL et QUAN. Dans ce chapitre, ce cadre est testé avec une revue des études utilisant les MM. Cette revue illustre plusieurs types de combinaisons avec des études en santé mondiale. Ensemble, ces deux chapitres contribuent à faire avancer les connaissances méthodologiques par (a) un appel à une meilleure description des MM dans les publications scientifiques, et (b) une conceptualisation testée proposant trois types d'intégration et neuf stratégies spécifiques, et toutes les combinaisons possibles des stratégies pour intégrer les phases, résultats et données QUAL et QUAN.

***

Mixed methods (MM) are increasingly popular. In performing MM, researchers integrate qualitative (QUAL) and quantitative (QUAN) methods with respect to worldviews, methodologies and designs, research questions, data collection and analysis techniques, and study results. Several integration strategies have been proposed, but their conceptualization is usually design-driven, or fragmented, or not empirically tested. This is challenging for planning and conducting MM studies, and for training graduate students. Based on the methodological literature, the previous chapter presented a conceptual framework including types of integration, practical strategies, and possible combinations of strategies for integrating QUAL and QUAN phases, results and data. In the present chapter, this framework is tested using a review of 2015 MM studies with a method-detailed description. This supports the framework, illustrates multiple types of combinations, and provides global health examples. These two companion chapters contribute to advance methodological knowledge on MM via (a) a call for better reporting MM studies in scientific articles, and (b) a tested conceptualisation comprising three types of integration and nine specific strategies, which explain current and future possibilities for combining strategies to integrate QUAL and QUAN phases, results, and data.

***

Los métodos mixtos (MM) son cada vez más populares. Al utilizar los métodos mixtos, los investigadores integran métodos cualitativos (QUAL) y cuantitativos (QUAN) con respecto a las cosmovisiones, metodologías y diseños, preguntas de investigación, técnicas de recopilación y análisis de datos y resultados de estudios. Se han propuesto varias estrategias de integración, pero su conceptualización suele estar impulsada por el diseño, o ser fragmentada, o no haber sido probada empíricamente. Esto es un reto para la planificación y realización de estudios que utilizan métodos mixtos y para la formación de estudiantes de postgrado. Basado en la literatura metodológica, este capítulo presenta un marco conceptual que incluye tipos de integración, estrategias prácticas y posibles combinaciones de estrategias para integrar las fases, resultados y datos de QUAL y QUAN. En el siguiente capítulo, este marco se prueba utilizando una revisión de los estudios que utilizan métodos mixtos de 2015 con una descripción detallada del método. Esto apoya el marco, ilustra múltiples tipos de combinaciones y proporciona ejemplos de salud mundial. Estos dos capítulos complementarios contribuyen a avanzar en el conocimiento metodológico de los métodos mixtos a través de (a) un llamado a informar mejor sobre los estudios que utilizan métodos mixtos en artículos científicos, y (b) una conceptualización probada que comprende tres tipos de integración y nueve estrategias específicas, que explican las posibilidades actuales y futuras de combinar estrategias para integrar las fases, los resultados y los datos de QUAL y QUAN.

***

Pierre Pluye est professeur titulaire au Département de médecine de famille de l'Université McGill, chercheur-boursier senior du Fonds de recherche du Québec Santé (FRQS), et directeur de la composante Développement méthodologique à l'Unité SOUTIEN SRAP du Québec. Il est membre associé de l'École des sciences de l'information de l'Université McGill. Il est spécialiste en médecine familiale et en santé publique. Il détient un doctorat en promotion de la santé (Université de Montréal). Il a obtenu des bourses de carrière des Instituts de recherche en santé du Canada (IRSC) et du FRQS. En 2017, il a été élu membre de l'Académie canadienne des sciences de la santé, et a reçu le prix du « Chercheur de l'année » du Collège des médecins de famille du Canada. Il possède une expertise en méthodes mixtes. Ses recherches étudient les effets des informations issues de ressources électroniques sur les professionnels, les patients et le public

Enrique García Bengoechea est chercheur dans la faculté d'éducation et sciences de la santé de l'université de Limerick (Irlande). Ses intérêts principaux de recherche sont la promotion de l'activité physique et la santé dans le contexte communautaire et le développement et socialisation des jeunes dans le sport. Préalablement, Enrique a travaillé comme chercheur dans le « Alberta Centre for Active Living », a enseigné aux universités McGill et Western Sydney, et a été conseiller en méthodes mixtes de recherche dans la plateforme de développements méthodologiques de l'Unité Soutien SRAP du Québec. Il a été aussi chercheur invité à l'institut du sport, exercice et la vie active (ISEAL) de l'université Victoria à Melbourne. Enrique est couramment membre de l'équipe de recherche de l'étude BuiltEnvironmentand Active Transport to School (BEATS), basée al'université d'Otago. Il est aussi affilié au groupe de recherche Éducation physique et promotion de l'activité physique de l'université de Zaragoza.

David Li Tang est agent de développement des services au Conseil national de recherches du Canada, et professeur adjoint au Département de médecine familiale de l'Université McGill. Son domaine de travail est axé sur l'innovation. David et ses collègues conçoivent et développent notamment des infrastructures et des méthodologies pour stimuler l'innovation au niveau collectif le plus large, impliquant des agences gouvernementales, des universités, des industries, des associations professionnelles, ainsi que des chercheuses et chercheurs individuels. Ils ont un impact sur les systèmes canadiens d'innovation et de soins de santé. La recherche multidisciplinaire de David couvre les sciences sociales, l'ingénierie et les sciences de l'information.

Vera Granikov est une spécialiste de l'information intégrée dans le Département de médecine de famille de l'Université McGill. Elle fournit des services informationnels adaptés aux chercheuses et chercheurs et aux cliniciens en soins primaires et coordonne le système de veille collaborative financé par l'Unité SOUTIEN SRAP du Québec. Vera poursuit actuellement des études de doctorat à l'École des sciences de l'information de l'Université McGillet détient une bourse au doctorat de Fonds de recherche du Québec – Société et culture (FRQSC).

Citation

Pierre Pluye, Enrique Garcìa Bengoechea, David Li Tang, Vera Granikov. (2019). La pratique de l'intégration en méthodes mixtes. Les multiples combinaisons des stratégies d'intégration. In *Évaluation des interventions de santé mondiale. Méthodes avancées*. Sous la direction de Valéry Ridde et Christian Dagenais, pp. 213-238. Québec : Éditions science et bien commun et Marseille : IRD Éditions.

PARTIE IV

L'ÉVALUATION DE L'EFFICACITÉ ET DE L'EFFICIENCE

9. Les méthodes quasi-expérimentales

L'effet de l'âge légal minimum sur la consommation d'alcool chez les jeunes aux États-Unis

TARIK BENMARHNIA ET DANIEL FULLER

Les méthodes quasi-expérimentales

Définition de la méthode

Le terme « quasi-expérimental » fait référence à une « expérimentation avec une intervention donnée, des résultats de santé mesurés et au moins deux unités expérimentales (groupe recevant l'intervention et groupe contrôle), mais sans recours à la randomisation pour créer les conditions d'inférence causale » (Cook, Campbell *et al.* 1979)

Forces de la méthode

Les méthodes quasi-expérimentales constituent un excellent devis d'étude alternatif aux essais contrôlés randomisés lorsque ces derniers sont difficiles ou impossibles à mener. En effet, ces méthodes permettent, en se basant sur des situations observées où il n'y a pas eu de répartition délibérée et aléatoire des groupes d'intervention et de contrôle, de se baser également sur la notion de contrefactuel pour estimer des effets de nature causale

Défis de la méthode

Chaque méthode quasi-expérimentale nécessite de respecter des hypothèses de travail supplémentaires. Parfois ces hypothèses sont impossibles à vérifier empiriquement. Les effets qui sont estimés peuvent varier en ce qui concerne les populations concernées et tout comme pour les essais contrôlés randomisés, il est important de distinguer validité interne et validité externe.

L'évaluation des effets des politiques publiques sur la santé des populations est un domaine qui s'est particulièrement développé dans les dernières années. L'évaluation des effets se distingue d'autres types d'évaluations telles que l'évaluation d'implantation (Sabatier et Mazmanian 1980; Belaid et Ridde 2012) ou l'évaluation économique (Brousselle et Lessard 2011; Drummond, Sculpher *et al.* 2015, Brousselle, Benmarhnia *et al.* 2016). Évaluer les effets des politiques publiques sur la santé des populations est essentiel pour plusieurs raisons (Champagne, Contandriopoulos *et al.* 2009; Macintyre 2010). Cela permet i) de s'assurer que ces politiques contribuent à améliorer l'état de santé de la population, ce qui constitue un des mandats de la santé publique, ii) de s'assurer que des objectifs d'un programme de

santé publique sont bel et bien atteints, iii) d'expérimenter de nouvelles interventions et de décider quant au maintien, à la modification ou à l'arrêt d'interventions existantes et iv) de s'assurer que les ressources publiques sont utilisées de manière efficace. Dit simplement, l'objectif principal de ce type d'évaluation est d'attribuer des changements à une intervention en particulier quant à un indicateur de santé spécifique. Autrement dit, il est question d'attribuer un effet causal à une intervention sur un indicateur de santé dans un contexte particulier.

La notion d'inférence causale est centrale dans de nombreuses disciplines aujourd'hui, incluant l'économétrie, la psychologie sociale ou l'épidémiologie. L'un des modèles les plus souvent utilisés pour identifier des effets de nature causale est basé sur la notion de contrefactuel (Susser 2001; Pearl 2009; Naimi et Kaufman 2015). Par exemple, pour évaluer si des changements de résultats de santé sont directement attribuables à une intervention, on cherchera à les comparer au sein d'un groupe observé (que l'on appellera factuel) à un groupe hypothétique semblable ou interchangeable (que l'on appellera contre-factuel) qui ne se distingue que par une seule caractéristique, à savoir le fait de ne pas avoir reçu l'intervention. Ainsi, on mesurera l'effet causal par le contraste entre le groupe factuel et le groupe contre-factuel. Le *Problème fondamental de la causalité* (Holland 1986; Kaufman 2007) est que, par définition, ce groupe contre-factuel n'est pas directement observable. Ainsi, la randomisation a été proposée comme solution pour pallier ce problème, tout en permettant de maintenir l'interchangeabilité et la substitution entre les groupes exposés à l'intervention ou non, et permettant de s'assurer de ne pas enfreindre des hypothèses permettant l'inférence causale, de façon à ce que l'intervention ne puisse avoir un effet sur les résultats de santé que du groupe factuel uniquement, par exemple. La randomisation permet notamment de pallier l'enjeu capital qu'est la prise en compte des facteurs de confusion mesurés ou non (Bonell, Hargreaves *et al.* 2011). C'est ainsi que les essais contrôlés randomisés (ECR) ont été introduits et utilisés en médecine et en sciences sociales : un groupe est assigné aléatoirement à l'intervention (ou au traitement) et un groupe ne l'est pas. Cette répartition aléatoire permet ainsi de s'assurer qu'il n'y a pas de différence entre les deux groupes avant l'intervention, autrement dit qu'il n'y a pas de biais de confusion, mesurés ou non. Cependant, ce type d'expérimentation, où l'on choisit délibérément d'intervenir sur un groupe et non sur un autre (le groupe contrôle) n'est pas

faisable dans de nombreuses situations, que ce soit pour des raisons de coût ou des raisons éthiques (Royall 1991; Hawe, Shiell *et al.* 2004; Moore et Moore 2011; Petticrew, Chalabi *et al.* 2011).

Les « expérimentations naturelles » constituent un excellent devis d'étude alternatif aux ECR, car elles permettent, en se basant sur des situations observées et sans répartition délibérée et aléatoire des groupes d'intervention et de contrôle, de se baser également sur la notion de contre-factuel pour estimer des effets de nature causale (Bor 2016). Pour cela, les méthodes quasi-expérimentales (MQE) ont été développées pour concevoir des groupes contre-factuels en se basant uniquement sur des données observées. Le terme de « quasi-expérimental » fait référence à une « expérimentation avec une intervention donnée, des résultats de santé mesurés et deux unités expérimentales (groupe recevant l'intervention et groupe contrôle), mais sans recours à la randomisation pour créer les conditions d'inférence causale » (Cook, Campbell *et al.* 1979 : 14); Shadish et Cook 2009). Ainsi, les MQE permettent d'imiter les résultats que produirait un ECR, tout en ne servant que de situations existantes. Plusieurs MQE ont été développées pour évaluer les effets des politiques publiques et commencent à être utilisées en santé publique.

Périmètre du chapitre

Ce chapitre vise à présenter deux principales méthodes quasi-expérimentales (MQE) et leur application dans une étude qui vise à évaluer l'effet de l'âge légal minimum sur la consommation d'alcool chez les jeunes. Ces MQE permettent de prendre en compte les facteurs de confusion non mesurés qui peuvent avoir un impact important sur la validité des estimations. Des méthodes qui permettent de prendre en compte uniquement les facteurs de confusion mesurés, comme des méthodes d'appariement utilisant un score de propension par exemple, sont parfois décrites comme MQE (Shadish, Luellen *et al.* 2006; Austin 2011). Cependant, dans ce chapitre, nous ne les aborderons pas. Nous nous focaliserons plutôt sur deux types de MQE : la méthode de « différence dans les différences » (DD), ainsi qu'un type particulier de variable instrumentale, à savoir l'approche de « régression avec discontinuité » (RD).

Les variables instrumentales pour évaluer l'effet des politiques publiques

sont couramment utilisées en économétrie (Newhouse et McClellan 1998; Hernán et Robins 2006) où, par ailleurs, une terminologie distincte est utilisée. Dit simplement, une variable instrumentale est une variable qui sera fortement corrélée à la probabilité de bénéficier d'une intervention, mais qui ne sera aucunement liée à l'issue de santé d'intérêt. Ainsi, on pourra utiliser cette variable instrumentale pour pallier les potentiels biais de confusions (mesurés ou non) dans la relation entre la probabilité de bénéficier de l'intervention et d'éventuels changements de l'issue de santé d'intérêt (voir Angrist et Pischke 2008 et 2014). Des variables géographiques (frontières, altitude, pluviométrie…) ou génétiques pourront notamment être utilisées comme variables instrumentales. Nous n'aborderons pas le cas général des variables instrumentales dans ce chapitre. De même, nous n'aborderons pas les analyses de séries temporelles interrompues comme MQE pour évaluer les effets des politiques publiques (voir Fuller, Sahlqvist *et al.* 2012; Benmarhnia, Zunzunegui *et al.* 2014; Kontopantelis, Doran *et al.* 2015).

L'objectif général de ce chapitre est de décrire deux MQE, de les appliquer empiriquement à un exemple basé sur une simulation et de fournir tous les éléments, incluant les hypothèses préalables à vérifier, les différentes étapes à mener et la syntaxe pour mener les analyses avec les logiciels *R* et *Stata*. Ceci permettra aux lecteurs et lectrices de reproduire ce type d'analyses dans leur contexte d'évaluation.

> Il est possible de retrouver en ligne toutes ces ressources informatiques, ainsi que toute la syntaxe permettant de les appliquer à d'autres questions de recherche, à l'adresse https://github.com/walkabilly/methode_evaluation. Il faut en profiter!

Présentation du contexte d'évaluation : les interventions pour réduire la consommation d'alcool chez les jeunes

Dans ce chapitre, nous prendrons comme exemple de politique publique

le changement de l'âge légal minimum pour consommer de l'alcool afin de réduire la consommation chez les jeunes.

La consommation d'alcool est un contributeur important au fardeau de morbidité et mortalité chez les adolescent-e-s et jeunes adultes dans de nombreux pays (Rehm, Gmel *et al.* 2003). Il existe de nombreuses interventions visant à prévenir la consommation d'alcool chez les jeunes. Depuis de nombreuses décennies, plusieurs pays ont mis en place un âge légal minimum pour la consommation d'alcool (Carpenter et Dobkin 2011). Malgré ce cadre légal et de nombreuses autres interventions de santé publique visant à réduire la consommation d'alcool chez les jeunes, l'alcool continue d'être un contributeur important à la mortalité. Ainsi, plusieurs pays ont décidé d'augmenter cet âge légal de consommation d'alcool avec pour objectif de réduire la consommation d'alcool et prévenir les accidents attribuables à la consommation. Quelques études ayant évalué ce type de politique publique avec des MQE de type DD ou RD ont été publiées ces dernières années (Donald et Lang 2007; Carpenter et Dobkin 2009; Yörük et Yörük 2012). Ici, en nous basant sur des données simulées, nous présenterons comment appliquer deux approches permettant d'estimer les effets de ce type de politique publique sur la consommation d'alcool chez les jeunes.

Présentation et application des méthodes proposées

Méthode de « différence dans les différences »

Présentation générale de la méthode DD

La méthode dite de différence dans les différences (DD) se sert d'un ou plusieurs groupes contrôle pour approximer la trajectoire d'un groupe contre-factuel n'ayant pas reçu d'intervention qui serait interchangeable avec le groupe ayant reçu l'intervention. Par exemple, ce ou ces groupe(s) contrôle pourront être des pays ou régions n'ayant jamais reçu ce type d'intervention ou l'ayant reçu à un moment différent du groupe d'intervention considéré. L'hypothèse principale à considérer dans ce type d'approche est qu'en l'absence de l'intervention, les tendances du groupe contrôle sont une bonne approximation des tendances du groupe qui

recevra l'intervention (Angrist et Pischke 2008; Dimick et Ryan 2014). Autrement dit, si l'intervention n'avait pas eu lieu, les tendances entre les deux groupes vis-à-vis de l'indicateur d'intérêt resteraient constantes.

Considérons schématiquement deux États dans le même pays, à quatre temps d'observation (2009 [t0], 2010 [t1], 2011 [t2], 2012 [t3]), un État implantant l'intervention en 2010 (t1) et l'autre État ne faisant rien. En assumant et/ou en vérifiant que les deux États sont interchangeables, on commencera par estimer la première différence qui sera par exemple, la différence de prévalence de consommation d'alcool chez les jeunes entre les deux États à toutes les périodes de temps confondues. Ensuite on estimera cette même différence en contrôlant l'effet du temps et de l'État. L'hypothèse nulle (l'intervention n'a pas d'effet sur la consommation d'alcool chez les jeunes) impliquerait que ces deux différences demeurent égales dans le temps. Ainsi, si les deux États sont interchangeables et que l'hypothèse de DD est respectée, alors toute différence entre ces deux différences pourra être interprétée comme l'effet de l'intervention sur la consommation d'alcool chez les jeunes dans l'État ayant reçu l'intervention.

Description des données

L'objectif de cette étude est d'estimer l'impact sur la consommation d'alcool d'une politique publique organisée en 2011 et ayant augmenté l'âge légal pour la consommation d'alcool de 18 à 19 ans dans 10 États du pays.

Afin de présenter l'application de la méthode de différences dans les différences, nous avons créé une base de données simulée qui contient des données semblables à un recensement national. Nous présumons 20 000 répondants âgés de 16 à 25 ans échantillonnés dans 20 États. La probabilité de sélection chez les répondant-e-s est égale pour chaque État. Nous présumons une probabilité égale de sélection pour éviter le besoin de pondération. Cependant, il est tout à fait possible d'inclure les pondérations dans l'application de cette méthode.

Le recensement contient les variables suivantes :

- *étatID* : l'identificateur unique pour les États.
- *individuID* : l'identificateur unique pour les répondant-e-s.

- *connaissance* : une variable individuelle de connaissance au sujet de l'alcool. Une augmentation de la valeur indique une connaissance plus importante au sujet de l'impact négatif de l'alcool.
- *effetTemps* : indique l'année avec des valeurs entre 2009 et 2012 inclusivement.
- *loi* : variable binaire qui indique si l'état a adopté ou non la loi augmentant l'âge de consommation de 18 à 19 ans (c'est-à-dire la politique publique).
- *âge* : une variable qui indique l'âge du répondant ou de la répondante.
- *consParSem* : nombre de consommation d'alcool par semaine du répondant ou de la répondante.

Ainsi, nous avons les données nécessaires pour faire une analyse de différences dans les différences. La base de données simulée présentée ici est typique pour une analyse de DD. Celle-ci inclut des données sur plusieurs années (deux au minimum), différents États, une politique implantée de manière différentielle à différentes années entre les États. Ces conditions sont nécessaires afin de construire le contre-factuel et mener l'analyse de DD.

Construire un contre-factuel

Le contre-factuel est représenté par une condition inobservable où la politique publique n'est pas implantée. Dans le cas où le contre-factuel est valide, nous ne présumons aucune modification de la consommation d'alcool dans les États où la politique n'a pas été implantée, donc pas d'effet de la politique publique. Nous avons inclus les ajustements par temps, État, âge et connaissances des effets négatifs de l'alcool sur la santé, afin de créer le contrefactuel le plus plausible possible. Afin que l'hypothèse principale de la méthode DD soit validée, nous devons aussi présumer que la pente de consommation par semaine dans le temps suit la même trajectoire dans les États d'intervention et de contrôle (voir Figure 1).

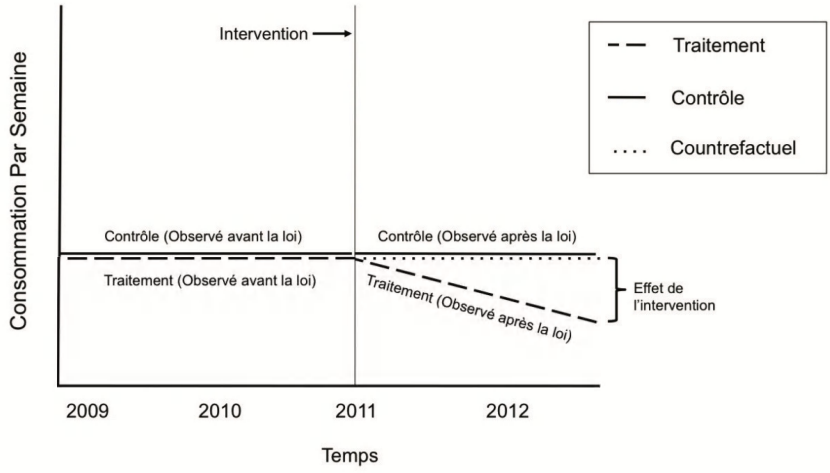


Figure 1. Représentation graphique de la méthode de différence dans les différences (DD)

Estimer l'effet de la politique publique

Avant de poursuivre l'estimation de l'effet de la politique publique, les procédures standard de nettoyage de données sont requises. Nos données simulées ne contiennent ni données manquantes ni données aberrantes. Étant donnée la distribution de notre variable d'intérêt (nombre de consommation d'alcool par semaine), nous allons utiliser un modèle de type binomial négatif (pour plus de détails, voir Hilbe 2011).

Afin d'estimer l'effet de la loi sur la consommation d'alcool en utilisant la méthode de DD, nous allons procéder en quatre étapes.

Étape 1 : La première étape consiste à inclure la variable d'intervention (loi) qui indique si un État a adopté ou non la loi augmentant l'âge de consommation d'alcool de 18 et 19 ans. Le Modèle 1 (Tableau 1) présente le résultat de l'inclusion de variable loi dans le modèle. L'effet de la loi est une réduction d'environ une unité de consommation d'alcool par personne chez les 16-25 ans (Beta : -1,11, IC 95 % -1.08 à -1,13). La variable loi est une comparaison brute entre le groupe traitement et le groupe contrôle, celle-ci ne contrôle pas deux facteurs importants, soit l'effet du temps et les différences entre les États.

Étape 2 : Le Modèle 2 (Tableau 1) ajoute un effet fixe pour la variable temps. Pour cela, nous avons inclus une variable catégorielle pour chaque année. L'inclusion de cette variable contrôle l'effet séculaire de la consommation d'alcool dans le temps, ceci sans contrainte linéaire sur l'effet du temps. L'effet estimé pour la variable loi sur notre variable d'intérêt reste semblable à l'étape précédente avec une réduction d'environ une unité de consommation d'alcool par personnes chez les 16-25 ans (Beta : -1,08, IC 95 % -1.08 à -1,03).

Étape 3 : Le Modèle 3 (Tableau 1) ajoute un effet fixe pour les États. Nous avons inclus une variable catégorielle pour chaque État. L'inclusion de cette variable permet de contrôler pour toute caractéristique spécifique au niveau de l'État observable et non observable qui est invariant dans le temps. L'inclusion des effets fixes pour le temps (cf. étape 2) et l'État nous permet d'estimer l'effet intra-État avant et après la loi entre les États avec traitement, ceux qui organisent la loi, et les États contrôle, ceux qui ne l'organisent pas. Dans le Modèle 3, l'effet estimé pour la variable loi sur notre variable d'intérêt reste statistiquement significatif, mais l'ampleur de l'association est réduite (-0,32, IC 95 % -0.39 à -0,24).

Étape 4 : Dans le Modèle 4 (Tableau 1), nous ajoutons les deux variables individuelles, le niveau de connaissance et l'âge des individus. L'inclusion de ces variables ne change pas notre estimation de l'effet de la loi sur la consommation (-0,32, IC 95 % -0.39 à -0,24).

Conclusion

Nous avons pu estimer, en utilisant la méthode de DD, l'effet d'une politique publique visant à changer l'âge légal sur la consommation d'alcool chez les 16-25 ans. Nous avons trouvé que cette politique publique contribuait à réduire la consommation d'alcool dans le groupe d'âge visé.

Ainsi, en utilisant les contrastes entre les États quant à la date d'implantation de cette politique publique, nous avons pu construire des contre-factuels qui nous ont permis de comparer les changements observés vis-à-vis de la consommation d'alcool chez les 16-25 ans à ce qui se serait passé sans l'implantation de cette politique.

Tableau 1. Résultats issus des modèles de Différence dans les Différences

	Model 1 Beta (IC 95 %)	**Model 2 Beta (IC 95%)**	**Model 3 Beta (IC 95%)**	**Model 4 Beta (IC 95%)**
Intercept	2.23 (2,21 à 2.24)	2.21 (2,19 à 2.24)	2.12 (2,07 à 2.18)	2.14 (2,02 à 2.25)
Loi				
Non	1	1	1	
Oui	-1,11 (-1,08 à -1,13)	-1,11 (-1,13 à -1,08)	-0,32 (-0,39 à -0,24)	-0,32 (-0,39 à -0,24)
Années				
2009		1	1	
2010		0.04 (0,00 à 0.07)	0.03 (-0,01 à 0.06)	0.03 (-0,01 à 0.06)
2011		0.04 (0,00 à 0.07)	0.04 (0,00 à 0.07)	0.04 (0,00 à 0.07)
2012		-0.02 (-0,06 à 0.01)	-0,04 (-0,07 à 0.00)	-0,04 (-0,07 à 0.00)
Connaissance				0.00 (-0,01 à 0.00)
Âge				0.00 (-0,01 à 0.00)

* Il est commun de ne pas présenter les résultats de l'effet fixe pour les États. Nous ne représentons donc pas les coefficients pour chaque État. L'État 20 est exclu du modèle à cause de la colinéarité avec la loi et le temps.

La méthode de « régression avec discontinuité »

Présentation générale de la méthode de RD

La méthode dite de régression avec discontinuité (RD) peut être utilisée lorsqu'une intervention se base sur un seuil précis d'une variable continue pour déterminer l'éligibilité quant à cette intervention. Ce seuil peut être

l'âge (comme dans notre exemple), un niveau de revenu (Andalón 2011) ou des marqueurs biologiques tels que les CD4, cellules du système immunitaire impactées par le VIH (Bor, Moscoe *et al.* 2014) par exemple. Tous les individus d'un côté du seuil recevront l'intervention tandis que tous ceux de l'autre côté de ce seuil ne la recevront pas. L'idée générale de cette approche est de considérer comme interchangeables les individus très proches d'une part et de l'autre de ce seuil, considérant le choix du seuil à sa valeur précise comme aléatoire. Le groupe contrefactuel est ainsi construit à partir d'une situation dans laquelle il n'y aurait aucune discontinuité au niveau du seuil d'éligibilité de l'intervention. Ainsi, il sera possible d'inférer qu'une différence dans les issues de santé entre les deux groupes sera la cause de la mise en place de ce seuil. Plusieurs étapes sont nécessaires pour mener à bien une analyse de type RD afin de vérifier plusieurs hypothèses quant à l'interchangeabilité entre les deux groupes, et le choix des modèles de régressions. Ces sept étapes sont détaillées ci-dessous.

Description des données

L'objectif de notre étude est d'estimer l'impact de l'âge légal sur la consommation d'alcool chez les jeunes. Afin de présenter l'application de la méthode de régression avec discontinuité, nous avons créé une base de données simulée. La base de données simulée contient des données pour 1 000 personnes âgées de 19 à 23 ans. De la même manière que pour l'analyse DD, aucune pondération n'a été incluse dans les analyses ici, mais cela demeure toutefois possible.

Nous avons une variable d'intérêt qui est la consommation d'alcool. Nous avons des données pour 1 000 personnes âgées de 19 à 23 ans.

La base de données contient les cinq variables suivantes :

- id : l'identificateur unique pour les répondant-e-s.
- âge : une variable qui indique l'âge du répondant ou de la répondante.
- propAlcJour : nombre de consommation d'alcool par jour du répondant ou de la répondante.
- fruit : consommation de fruits et légumes par semaine.
- cigarette : consommation de cigarettes par semaine.

Application de la méthode RD en sept étapes

1. Définir le seuil

Les données requièrent un seuil dans la probabilité de consommation par rapport à une variable continue d'attribution. Dans notre exemple, l'âge est la variable d'attribution. Ici, l'âge de majorité est 21 ans. Donc, avant 21 ans la consommation d'alcool n'est pas légale et après 21 ans elle est légale. Ceci constitue une rupture nette à l'âge de 21 ans.

2. Vérifier qu'il n'y a pas de biais ou erreur de mesure pour la variable d'intérêt par rapport au seuil

Un biais ou erreur de mesure pour la variable d'intérêt par rapport au seuil créera un biais systématique dans les estimations par la méthode RD. Par exemple, si la consommation d'alcool est sous-estimée chez les participant-e-s âgés de moins de 21 ans, soit parce que les répondant-e-s ne donnent pas des réponses réelles (ou refusent de répondre) puisque le comportement est illégal, l'estimé RD est biaisé de façon systématique (Imbens and Lemieux 2008). Dans notre cas, nous présumons qu'il n'y a pas de biais ou erreur de mesure pour la variable d'intérêt.

3. Vérifier que la variable d'attribution est continue et mesurée précisément

Dans la troisième étape, nous visualisons un histogramme, un graphique « Q-Q », et nous menons un test *Shapiro-Wilk* afin de nous assurer que la variable d'attribution continue est uniformément distribuée. Une variable d'attribution qui n'est pas continue posera problème dans le contexte de la méthode régression avec discontinuité, puisqu'il est possible que plusieurs seuils existent. Ainsi, si la variable d'attribution et plusieurs seuils existent, il n'est pas possible d'identifier un effet causal puisqu'un effet est possible à chaque seuil. La variable âge dans notre échantillon est uniforme.

4. Visualiser la variable d'attribution par la variable d'intérêt

Dans la quatrième étape, nous visualisons la variable d'attribution sur l'axe X avec les variables d'intérêts sur l'axe Y. La figure 2 présente ces relations. Nous présentons les données complètes avant et après l'âge de 21 ans en utilisant des options de graphique particulières. Le « *smoothing* » loess est le graphique créé par défaut par la commande rdrobust avec « *binning* » (*mimicking variance evenly-spaced method using spacings*

estimators) et « *smoothing* » avec un polynomial du 4iéme ordre. Il est recommandé de présenter un graphique avec les données « *binned* » (Calonico, Cattaneo *et al.* 2015).

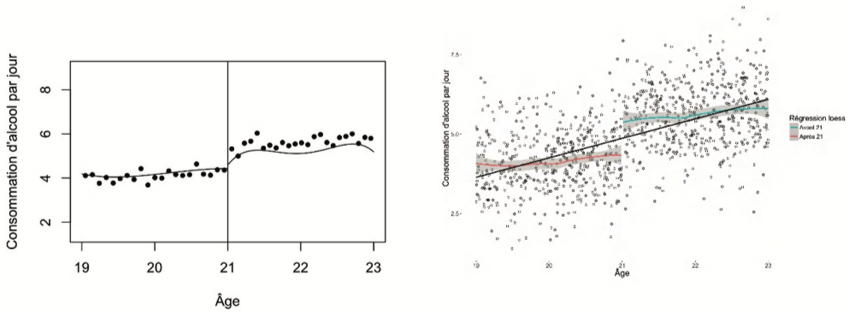


Figure 2. Représentation graphique de la méthode de Régression avec Discontinuité

5. Comparaison de variable d'intérêt et co-variables

Dans la cinquième étape, nous voulons vérifier que les deux groupes sont interchangeables, sauf pour la variable sanitaire d'intérêt. Nous comparons la moyenne et l'écart-type de la variable d'intérêt avant et après 21 ans. Nous examinons aussi ceci pour deux co-variables que sont la consommation de fruits et légumes et la consommation de cigarettes (tableau 2). Étant donnée la prémisse que la politique influence seulement la consommation d'alcool, la politique ne devrait pas influencer d'autres comportements. L'idée de cette étape est de s'assurer qu'à part pour la consommation d'alcool, il n'y a pas d'autre discontinuité au même seuil (c'est-à-dire à 21 ans) pour d'autres variables dans l'échantillon.

Tableau 2. Résultats issus des modèles de Régression avec Discontinuité

	Avant 21 ans (moyenne et SD)	**Après 21 ans (moyenne et SD)**
Alcool	4.2 (1,0)	5.6 (1,0)
Fruit	7.5 (4,5)	7.6 (4,3)
Cigarette	39.0 (22,0)	40.0 (23,0)

6. Sélectionner l'étendue de l'échantillon sélectionné

Une prémisse importante de la méthode RD est que les groupes d'un côté et de l'autre du seuil sont interchangeables. Cela signifie qu'ils doivent être similaires pour toute autre caractéristique que la variable d'attribution (l'âge dans notre exemple) et la variable sanitaire d'intérêt. Il y a plusieurs méthodes proposées pour sélectionner le « *bandwidth* ». Le « *bandwidth* » définit l'étendue de l'échantillon sélectionné pour la régression. Il est très important puisqu'il définit la plausibilité que les groupes d'un côté et de l'autre du seuil soient interchangeables. Le « *bandwidth* » est typiquement estimé avec une estimation par noyau. Ici, nous proposons deux méthodes de sélections, à savoir MSE [*Mean Square Error Optimal Bandwidth Selector*] et CER [*Coverage Error Rate Optimal Bandwidth Selector*]) (Imbens and Kalyanaraman 2011, Calonico, Cattaneo *et al.* 2015) et trois différentes méthodes d'estimation par noyau (triangulaire, Epanechnikov et uniforme). Les méthodes de sélections de *bandwidth* sont disponibles avec le programme rdrobust pour R (Calonico, Cattaneo *et al.* 2015) et Stata (Calonico, Cattaneo *et al.* 2014).

Le choix de *bandwidth* peut avoir un impact sur la taille et l'échantillon afin de respecter le fait que les deux groupes d'un côté et de l'autre du seuil sont interchangeables. Dans notre exemple, notre échantillon de 1 000 répondant-e-s n'est pas affecté par le choix de *bandwidth* avec 484 droites et 516 gauches répondants inclus dans l'analyse pour une sélection *bandwidth* IK MSE avec une estimation par noyau triangulaire ou une sélection *bandwidth* CT avec une estimation par noyau uniforme (voir tableau 3 pour les résultats complets).

Dans notre cas, nous allons choisir la méthode MSE avec l'estimation par noyau triangulaire pour les analyses, c'est-à-dire la méthode par défaut sous R avec *rdrobust*. Quel que soit le choix, il est nécessaire de faire des analyses de sensibilité afin de s'assurer que le résultat n'est pas affecté par le choix du modèle.

7. Estimer l'effet de la politique publique par des modèles de régression

D'abord, nous estimons l'effet de la politique avec une régression linéaire sur l'échantillon complet. L'estimation pour la régression linéaire est une augmentation de la consommation d'alcool de 0,6 par semaine (IC 95 % 0.6 à 0,7). Avec l'échantillon complet, la prémisse que les groupes d'un côté et de l'autre du seuil sont interchangeables est moins plausible. Afin d'estimer

un effet plausible, nous estimons la régression avec discontinuité avec la méthode MSE et l'estimation par noyau triangulaire. La régression inclut désormais 484 répondants à la gauche (avant 21 ans) et 516 répondants à la droite (après 21 ans) (tableau 4). La régression est estimée avec une régression locale linéaire et correction locale quadratique. L'estimation n'est pas influencée par la méthode de sélection du « *bandwidth* », ce qui indique que le résultat est robuste.

Tableau 3. Estimation de l'effet de la consommation d'alcool avec RD

	Bandwidth MSE	**Bandwidth CER**
	Beta (IC 95 %)	Beta (IC 95 %)
Conventional	0.88 (0,2 à 1,5)	0.80 (0,0 à 1,6)
Bias-Corrected	0.80 (0,2 à 1,4)	0.75 (-0,0 à 1,5)
Robuste	0.80 (0,0 à 1,6)	0.75 (-0,1 à 1,6)

Les résultats montrent que le fait d'appartenir au groupe d'âge ayant le droit de consommer de l'alcool augmente d'environ une unité de consommation d'alcool par semaine (IC 95 % 0,2 à 1.5). Les analyses de sensibilité (CER avec correction pour le biais et robuste) démontrent qu'il est possible que l'estimation soit sensible au choix de modèle. Le chercheur ou la chercheuse doit être prudent-e dans l'interprétation et la communication des résultats.

Conclusion

Nous avons pu estimer, en utilisant la méthode de RD, l'effet de l'âge légal minimum de consommation d'alcool sur la consommation d'alcool chez les 19-23 ans. Nous avons trouvé que la consommation d'alcool augmente d'une unité de consommation par semaine quand l'on passe de 20 à 21 ans, ce qui montre que cette politique publique a un effet sur la consommation d'alcool chez les 19-23 ans.

En utilisant les contrastes entre les 19 à 20 et les 21 à 23 ans, nous avons pu construire un contre-factuel qui nous a permis de comparer les changements observés vis-à-vis de la consommation d'alcool chez les

19-23 ans avec ce qui serait passé sans la mise en place de cette politique publique.

Discussion

Dans ce chapitre, nous avons présenté deux méthodes quasi-expérimentales dans le contexte d'évaluation des effets de politiques publiques sur la santé. Nous avons montré que ces approches pouvaient constituer de bonnes solutions lorsque des essais contrôlés randomisés sont impossibles à mener. Nous avons principalement présenté deux approches, soit la méthode de différence dans les différences et l'approche de régression avec discontinuité. Ces deux approches permettent d'exploiter des situations d'expérimentations naturelles différentes. Nous avons, dans un premier temps, utilisé pour la méthode de différence dans les différences la date d'implantation de l'intervention pour comparer les changements concernant des indicateurs de santé avec le groupe contrôle. Dans le cas de l'approche de régression avec discontinuité, nous avons plutôt exploité un caractère spécifique de l'intervention, à savoir l'âge légal de consommation d'alcool. Ces deux approches visent à répondre à la même question, mais la différence principale réside dans l'effet qui est estimé et la population à laquelle ces résultats peuvent être généralisés. Dans le premier cas, nous quantifions un effet moyen total (*Average Treatment Effect*) qui prend la population dans son ensemble (il est important de noter qu'il est aussi possible de quantifier l'effet sur des sous-groupes de la population). Dans le second cas, nous quantifions un effet moyen total local (*Local Average Treatment Effect*) au sein de la population incluse dans l'analyse (à savoir les personnes aux âges 19-23 ans) si bien que les résultats ne sont pas extrapolables aux autres groupes d'âge notamment.

L'ensemble des étapes nécessaires pour conduire ce type d'analyse ainsi que les hypothèses à vérifier ont été décrites. Ces approches, particulièrement développées dans le domaine de l'économétrie, ont beaucoup été utilisées pour évaluer l'efficacité d'interventions sur la santé dans des pays à bas et moyen revenus, notamment en ce qui concerne des programmes de transferts de fonds (conditionnels ou non) (Andalón 2011) ou de programme d'accès aux soins (Dimick et Ryan 2014). Il est important de préciser qu'il est possible de conduire ce type d'évaluation en utilisant des

données déjà collectées dans les pays à bas et moyen revenus, permettant ainsi la conduite de nombreuses recherches dans le futur.

Dans ce chapitre, nous avons abordé uniquement un type d'évaluation, à savoir l'évaluation des effets ou des impacts d'une politique publique, mobilisant la notion d'inférence causale. Inévitablement, ce type d'évaluation est complémentaire d'autres types, telles que les évaluations d'implantation qui mobilisent des méthodologies différentes. Enfin, nous avons abordé ici l'impact sanitaire de politiques publiques sur une population d'étude dans son ensemble. Pourtant, il est possible, voire systématique que les effets d'une politique publique ne soient pas distribués de manière homogène au sein de plusieurs sous-groupes d'une population. Cet aspect, à travers la notion d'équité, est ainsi abordé dans le chapitre suivant.

Références clés

Angrist, J. D. et Pischke, J.-S. (2014). *Mastering metrics : The Path from Cause to Effect*. Lieu d'édition : Princeton University Press.

Cet ouvrage est une excellente introduction aux méthodes quasi-expérimentales et autres approches couramment utilisées en économétrie et en épidémiologie. Il constitue une version plus accessible que l'ouvrage Mostly harmless econometrics: An empiricist's companion (Princeton University Press, 2008) par les mêmes auteurs, plus détaillé mais plus difficile d'accès.

Bor, J. (2016). Capitalizing on natural experiments to improve our understanding of population health. *American Journal of Public Health*, 106(8), 1388.

Cet article, très court et facile d'accès, constitue un très bon résumé de la notion d'expérimentations naturelles et comment leur exploitation peut constituer d'excellentes alternatives aux essais contrôlés randomisés.

Deaton, A. et Cartwright, N. (2017). Understanding and misunderstanding randomized controlled trials. *Social Science et Medicine*, *volume*(numéro), p.-p.

Cet article résume de nombreux travaux des dernières années visant à expliciter ce que sont et surtout ne sont pas les essais contrôlés randomisés. Cet article, et les nombreux commentaires l'accompagnant, sont à lire pour

comprendre les limites et les avantages de ce type d'approche qui est malheureusement trop souvent mal compris et mobilisé.

Références

Andalón, M. (2011). Oportunidades to reduce overweight and obesity in Mexico? *Health economics*, 20(S1), 1-18.

Angrist, J. D. et Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton : Princeton University Press.

Angrist, J. D. et Pischke, J.-S. (2014). *Mastering metrics : The path from cause to effect*. Princeton : Princeton University Press.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research*, 46(3), 399-424.

Belaid, L. et Ridde, V. (2012). An implementation evaluation of a policy aiming to improve financial access to maternal health care in Djibo district, Burkina Faso. *BMC pregnancy and childbirth*, 12(1), 1.

Benmarhnia, T., Zunzunegui, M.-V., Llácer, A. et Béland, F. (2014). Impact of the economic crisis on the health of older persons in Spain: research clues based on an analysis of mortality. SESPAS report 2014. *Gaceta Sanitaria*. 28, 137-141.

Bonell, C. P., Hargreaves, J., Cousens, S., Ross, D., Hayes, R., Petticrew, M. et Kirkwood, B. (2011). Alternatives to randomisation in the evaluation of public health interventions: design challenges and solutions. *Journal of Epidemiology and Community Health*, 65(7): 582-587.

Bor, J. (2016). Capitalizing on natural experiments to improve our understanding of population health. *American Journal of Public Health*, 106(8), 1388.

Bor, J., Moscoe, E., Mutevedzi, P., Newell, M.-L. et Bärnighausen, T. (2014). Regression discontinuity designs in epidemiology: causal inference without randomized trials. *Epidemiology*, 25(5), 729-737.

Brousselle, A., Benmarhnia, T. et Benhadj, L. (2016). What are the benefits and risks of using return on investment to defend public health programs?. *Preventive Medicine Reports*, 3, 135-138.

Brousselle, A. et Lessard, C. (2011). Economic evaluation to inform health care decision-making: promise, pitfalls and a proposal for an alternative path. *Social science et medicine*,72(6), 832-839.

Calonico, S., Cattaneo, M. D. et Farrell, M. H. (2015). On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association*, 113(522), 767-779.

Calonico, S., Cattaneo, M. D. et Titiunik, R. (2014). Robust data-driven inference in the regression-discontinuity design. *Stata Journal*, 14(4), 909-946.

Calonico, S., Cattaneo, M. D. et Titiunik, R. (2015). Optimal data-driven regression discontinuity plots. *Journal of the American Statistical Association*, 110(512), 1753-1769.

Calonico, S., Cattaneo, M. D. et Titiunik, R. (2015). rdrobust: An r package for robust nonparametric inference in regression-discontinuity designs. *R Journal*,7(1), 38-51.

Carpenter, C. et Dobkin, C. (2009). The effect of alcohol consumption on mortality: regression discontinuity evidence from the minimum drinking age. *American Economic Journal: Applied Economics*, 1(1), 164-182.

Carpenter, C. et Dobkin, C. (2011). The minimum legal drinking age and public health. *The Journal of Economic Perspectives*, 25(2), 133-156.

Champagne, F., Contandriopoulos, A.-P., Brousselle, A., Hartz, Z. et Denis, J.-L. (2009). L'évaluation dans le domaine de la santé : concepts et méthodes. In *L'évaluation : Concepts et méthodes*. Montréal : Presses de l'Université de Montréal.

Cook, T. D., Campbell, D. T. et Day, A. (1979). *Quasi-experimentation : Design et analysis issues for field settings*. Boston : Houghton Mifflin.

Dimick, J. B. et Ryan, A. M. (2014). Methods for evaluating changes in health care policy: the difference-in-differences approach. *Jama*, 312(22), 2401-2402.

Donald, S. G. et Lang, K. (2007). Inference with difference-in-differences and other panel data. *The review of Economics and Statistics*, 89(2), 221-233.

Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L. et Torrance, G. W. (2015). *Methods for the economic evaluation of health care programmes*. Lieu d'édition : Oxford University Press.

Fuller, D., Sahlqvist, S., Cummins, S. et Ogilvie, D. (2012). The impact of public transportation strikes on use of a bicycle share program in London: interrupted time series design. *Preventive medicine*, 54(1), 74-76.

Hawe, P., Shiell, A. et Riley, T. (2004). Complex interventions: how "out of control" can a randomised controlled trial be?. *Penelope*, 328, 1561-1563.

Hernán, M. A. et Robins, J. M. (2006). Instruments for causal inference: an epidemiologist's dream?. *Epidemiology*, 17(4), 360-372.

Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge : Cambridge University Press.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association*, 81(396), 945-960.

Imbens, G. et Kalyanaraman, K. (2011). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of economic studies*, 79(3), 933-959.

Imbens, G. W. et Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2), 615-635.

Kaufman, J. S. (2007). Making causal inferences about macrosocial factors as a basis for public health policies. *Macrosocial Determinants of Population Health*, *chapter* 17, 355-373.

Kontopantelis, E., Doran, T., Springate, D. A., Buchan, I. et Reeves, D. (2015). Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. *British Medical Journal*, 350, h2750.

Macintyre, S. (2010). Good intentions and received wisdom are not good enough: the need for controlled trials in public health. *Journal of Epidemiology and Community Health*, 65(7), 564-567.

Moore, L. et Moore, G. F. (2011). Public health evaluation: which designs work, for whom and under what circumstances?. *Journal of Epidemiology and Community Health*, 65(7), 596-597.

Naimi, A. I. et Kaufman, J. S. (2015). Counterfactual theory in social epidemiology: reconciling analysis and action for the social determinants of health. *Current Epidemiology Reports*, 2(1), 52-60.

Newhouse, J. P. et McClellan, M. (1998). Econometrics in outcomes research:

the use of instrumental variables. *Annual Review of Public Health*, 19(1), 17-34.

Pearl, J. (2009). *Causality*. Cambridge : Cambridge University Press.

Petticrew, M., Chalabi, Z. et Jones, D. R. (2011). To RCT or not to RCT: deciding when 'more evidence is needed' for public health policy and practice. *Journal of Epidemiology and Community Health*, 66(5), 391-396.

Rehm, J., Gmel, G., Sempos, C. T. et Trevisan, M. (2003). Alcohol-related morbidity and mortality. *Alcohol Research et Health*, 27(1), 39-51.

Royall, R. M. (1991). Ethics and statistics in randomized clinical trials. *Statistical Science*, 6(1), 52-62.

Sabatier, P. et Mazmanian, D. (1980). The implementation of public policy: A framework of analysis. *Policy Studies Journal*, 8(4), 538-560.

Shadish, W. R. et Cook, T. D. (2009). The renaissance of field experimentation in evaluating interventions. *Annual Review of Psychology*, 60, 607-629.

Shadish, W. R., Luellen, J. K. et Clark, M. (2006). Propensity Scores and Quasi-Experiments: A Testimony to the Practical Side of Lee Sechrest. *Dans R. R. Bootzin et P. E.* McKnight (dir.), *Strengthening research methodology: Psychological measurement and evaluation* (p. 143-157). Washington, DC : American Psychological Association.

Susser, M. (2001). Glossary: causality in public health science. *Journal of Epidemiology and Community Health*, 55(6), 376-378.

Yörük, C. E. et Yörük, B. K. (2012). The impact of drinking on psychological well-being: Evidence from minimum drinking age laws in the United States. *Social Science et Medicine*, 75(10), 1844-1854.

Résumé / Abstract

Les méthodes quasi-expérimentales (MQE) sont des alternatives aux essais contrôlés randomisés, qui permettent, en se basant sur le modèle contre-factuel, d'évaluer les effets d'interventions en exploitant des expérimentations naturelles. L'objectif de ce chapitre est d'introduire deux MQE, à savoir la me la méthode de Différence dans les Différences (DD) ainsi que l'approche de Régression avec Discontinuité (RD). Leur application afin d'évaluer l'effet de l'âge légal minimum sur la consommation d'alcool chez les

jeunes est ensuite présentée. Sont présentées toutes les étapes nécessaires et les codes statistiques (pour R et Stata) permettant de reproduire ce type d'analyse. Ces approches constituent de bonnes solutions lorsque des essais contrôlés randomisés sont impossibles ou difficiles à mener, permettant de s'assurer que des objectifs d'un programme de santé publique sont bel et bien atteints, de décider quant au maintien, à la modification ou à l'arrêt d'interventions existantes et de s'assurer que les ressources publiques sont utilisées de manière efficace.

***

Quasi-experimental methods have been developed to evaluate 'natural experiments,' an alternative design to randomized controlled trials that relies on the notion of the counterfactual to estimate causal effects. The objective of this chapter is to describe quasi-experimental methods, apply them empirically to an example, and provide all elements of the evaluation, allowing the reader to reproduce these analyses (in R and Stata) in their given context. The main objective of this type of policy evaluation is to attribute a causal effect of an intervention on a health indicator. We apply these methods to estimate the impact of a policy raising minimum legal drinking on alcohol consumption among young people. Two types of quasi-experimental methods are used in this evaluation: the Difference in Differences (DD) method and the discontinuous regression (RD) approach. We show that these approaches can be good solutions to evaluate policies when randomized controlled trials are impossible. These evaluation tools can be used to ensure that policies contribute to the improvement of population health, to inform and decide whether to maintain, modify or discontinue existing interventions and to ensure public resources are used effectively.

***

Se han desarrollado métodos cuasi-experimentales para evaluar los « experimentos naturales », un diseño alternativo a los ensayos controlados aleatorios que se basa en la noción de lo contrafactual para estimar los efectos causales. El objetivo de este capítulo es describir métodos cuasi-experimentales, aplicarlos empíricamente a un ejemplo y proporcionar todos los elementos de la evaluación, permitiendo al lector reproducir estos análisis (en R y Stata) en su contexto dado. El objetivo principal de este tipo de evaluación de políticas es atribuir un efecto causal de una intervención a

un indicador de salud. Aplicamos estos métodos para estimar el impacto de una política que aumenta el consumo mínimo legal de alcohol en el consumo de alcohol entre los jóvenes. En esta evaluación se utilizan dos tipos de métodos cuasiexperimentales: el método de la Diferencia de Diferencias (DD) y el enfoque de regresión discontinua (RD). Demostramos que estos enfoques pueden ser buenas soluciones para evaluar las políticas cuando los ensayos controlados aleatorios son imposibles. Estas herramientas de evaluación pueden utilizarse para garantizar que las políticas contribuyan a mejorar la salud de la población, para informar y decidir si se mantienen, modifican o interrumpen las intervenciones existentes y para garantizar que los recursos públicos se utilicen de manera eficaz.

***

Tarik Benmarhnia est Professeur Assistant d'épidémiologie à l'Université de Californie à San Diego au sein de l'Institut d'océanographie Scrippset de l'école de médecine. Il a complété son doctorat en épidémiologie à l'Université de Montréal et à l'Université Paris Sud. Il est également chercheur associé au Chili et en France et membre du comité éditorial de *Environmental Health Perspectives*. Sa recherche porte sur les impacts des événements extrêmes sur la santé avec un focus sur les populations vulnérables et les implications en termes de politiques publiques. Il développe également des approches méthodologiques pour évaluer l'effet des politiques environnementales sur la santé comme les mesures d'adaptation aux changements climatiques et mesures de contrôle de la pollution atmosphérique.

Daniel Fuller est titulaire d'une Chaire de recherche du Canada en activité physique populationnelle à l'Université Memorial de Saint-Jean de Terre-Neuve. Ses recherches portent sur les technologies portables afin d'étudier l'activité physique, les interventions en transports et l'iniquité dans les espaces urbains. Son travail méthodologique vise les méthodes d'expérimentation naturelle et l'apprentissage-machine. Dan est titulaire d'une maitrise en kinésiologie de l'Université de Saskatchewan, d'un doctorat en santé publique de l'Université de Montréal, et a effectué de la recherche postdoctorale au Département de santé communautaire et d'épidémiologie de l'Université de Saskatchewan. Daniel est co-directeur de l'équipe Neighbourhood Factors du Canadian Urban Environmental Health (CANUE) Research Consortium et chercheur principal de l'équipe INTERACT.

Citation

Tarik Benmarhnia et Daniel Fuller. (2019). Les méthodes quasi expérimentales. L'effet de l'âge légal minimum sur la consommation d'alcool chez les jeunes aux États-Unis. In *Évaluation des interventions de santé mondiale. Méthodes avancées*. Sous la direction de Valéry Ridde et Christian Dagenais, pp. 241-264. Québec : Éditions science et bien commun et Marseille : IRD Éditions.

10. Les essais randomisés en grappe

Un exemple en santé maternelle et infantile

ALEXANDRE DUMONT

Les essais randomisés en grappe

Définition de la méthode

Les essais randomisés en grappe permettent d'évaluer l'impact des interventions en santé mise en œuvre à l'échelle d'un groupe d'individus, tels que les patientes qui accouchent dans un même hôpital ou les habitants d'un même village.

Forces de la méthode

- permet d'éviter le biais de contamination par l'intervention des individus (ou cibles) de deux grappes différentes
- permet d'avouer les effets de l'intervention avec un niveau de preuve élevé
- les essais contrôlés en grappe sont le plus souvent des études expérimentales pragmatiques qui évaluent les interventions dans la vie réelle

Défis de la méthode

- l'intervention doit être la plus standardisée possible
- défis statistiques liés au fait que l'unité de randomisation et d'analyse n'est pas la même
- défis dans la sélection des grappes qui doivent être représentatives du système de santé où se déroule la recherche

La mise en œuvre des politiques de santé nécessite d'expérimenter régulièrement de nouvelles interventions adaptées aux contextes et censées faciliter l'implantation de bonnes pratiques ayant un fort impact sur la santé des populations. Dans la lutte contre la mortalité maternelle, la plupart des politiques de santé périnatale dans le monde sont basées sur l'accouchement assisté par du personnel qualifié et le traitement des

Maternité en RDC. Crédit « Action santé des femmes ».

complications obstétricales lorsqu'elles surviennent pendant le travail ou l'accouchement. Ces traitements sont connus, mais le défi pour la plupart des systèmes de santé en développement est de les mettre en œuvre en temps utile. De nombreuses initiatives ont été testées dans les pays à faibles ou moyens revenus (PFMR) pour mobiliser les communautés ou améliorer la filière de soins et les pratiques médicales afin que les femmes puissent accéder plus rapidement à des services obstétricaux de qualité.

Quelques évaluations expérimentales ou quasi-expérimentales ont permis de fournir des premières preuves sur l'efficacité et la mise en œuvre des interventions qui visent à réduire la mortalité maternelle. Mais ces informations sont encore fragmentaires et peu d'études d'impact de bonne qualité ont été réalisées dans les PFMR en comparaison avec les pays où les niveaux de ressources sont plus élevés. Il est pourtant nécessaire de produire et de diffuser plus de données probantes sur l'efficacité de ces interventions, qu'il s'agisse de programmes nationaux ou d'initiatives plus localisées. Ceci doit permettre aux décideuses et décideurs et aux responsables de programmes de santé maternelle de concevoir des politiques publiques plus efficaces.

Le niveau de preuve des études d'impact en population peut être comparé à celui des études cliniques pour les traitements médicamenteux (voir Tableau 1). Les études avec le niveau de preuve le plus élevé sont les essais contrôlés randomisés (ECR) dans lesquels la plupart des biais de sélection des participant-e-s sont contrôlés, permettant ainsi d'affirmer que les changements observés sont bien attribuables à l'intervention et non aux différences entre les individus ou les pratiques de soins. Il existe deux catégories d'ECR : les ECR individuels où les sujets sont tirés au sort individuellement pour participer à différents groupes d'intervention (randomisation individuelle) et les ECR en grappes dans lesquels on randomise des groupes de sujets (randomisation en grappe).

Tableau 1. Niveaux de preuve scientifique et grades des recommandations (Source : Hautes Autorités de Santé, France)

Niveau de preuve scientifique fourni par la littérature (études thérapeutiques)	**Grade des recommandations**
Niveau 1 (NP1) Essais comparatifs randomisés de forte puissance. Méta-analyse d'essais comparatifs randomisés Analyse de décision basée sur des études bien menées	Preuve scientifique établie A
Niveau 2 (NP2) Essais comparatifs randomisés de faible puissance. Études comparatives non randomisées bien menées Études de cohorte	Présomption scientifique B
Niveau 3 (NP3) Études cas-témoins	Faible niveau de preuve C

Les ECR individuels sont peu fréquents dans les PFMR pour trois raisons essentielles : (i) le coût de ces études est élevé, car elles nécessitent le plus souvent le recrutement de patient-e-s dans de multiples centres pendant plusieurs années; (ii) la plupart des interventions qu'il faudrait tester dans un contexte à faibles ressources ont déjà une efficacité connue dans les pays à ressources élevées; (iii) la randomisation de sujets n'est parfois pas possible pour évaluer des interventions dans la communauté pour des raisons logistiques, administratives, voire éthiques. Un autre inconvénient des ECR individuels est lié au fait que le recrutement des sujets est souvent très sélectif en raison des multiples critères d'inclusion et d'exclusion et les conditions dans lesquelles les interventions sont mises en œuvre sont trop standardisées. Les résultats obtenus par un ECR sont alors difficilement généralisables à d'autres contextes que celui dans lequel l'étude a été menée.

L'approche utilisée dans les essais randomisés en grappes est beaucoup plus pragmatique que l'ECR individuel (Tableau 2). L'intervention testée ne cible pas directement l'individu, mais une unité « sociale » indépendante (la famille, le village, le médecin, l'hôpital…) pour faire changer les comportements ou introduire de nouvelles pratiques dans la « vie réelle ».

Tableau 2. Essais randomisés individuels et en grappes

	ECR individuel	ECR en grappes
Unité d'intervention et de randomisation	Sujet	Groupe de sujets
Type d'intervention	Traitement	Programme d'intervention
Objectif	Améliorer l'état de santé	Changer les comportements
Risque de contamination	Élevé	Faible
Contexte d'étude	Très standardisé Milieu clinique	« Vie réelle » En population
Coût	Élevé	Élevé

Défis des essais contrôlés randomisés en grappes

Si les essais randomisés en grappes offrent plusieurs avantages dans les PFMR par rapport aux essais individuels, ils présentent quelques défis méthodologiques qui tiennent au fait que l'unité d'analyse est différente de l'unité de randomisation ou d'intervention (Figure 1).

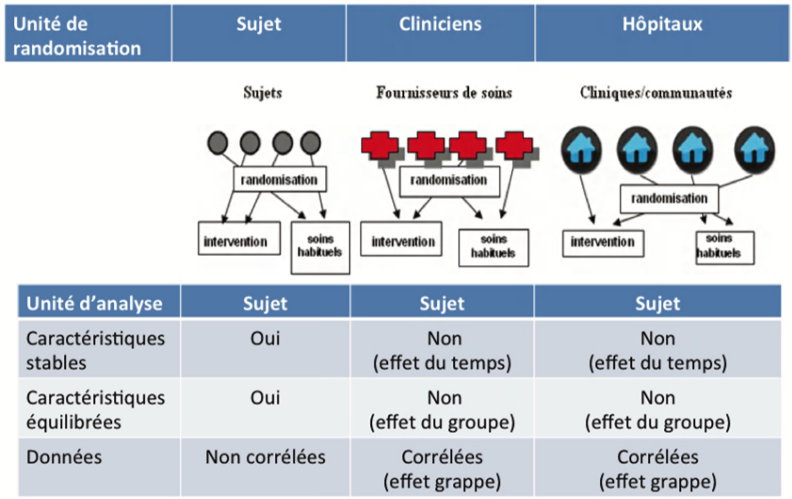


Figure 1. Implications des essais randomisés contrôlés en grappes

En randomisant des grappes, on se met en situation de recueillir des informations au niveau des individus dont les caractéristiques ne sont pas nécessairement stables au cours du temps, ni équilibrées entre les bras de l'essai. De plus, ces données sont corrélées au sein de chaque grappe. Cette situation conduit à trois conséquences importantes à considérer pour l'analyse des résultats d'un essai en grappe.

Premièrement, les caractéristiques des individus au sein des grappes, mais aussi la performance globale du système de santé, peuvent évoluer pendant la période d'étude. C'est ce qui explique les changements observés des indicateurs de santé dans les grappes où aucune intervention extérieure n'est mise en œuvre (groupe contrôle) et que nous nommerons « l'effet du temps ». Par exemple, le ratio de mortalité maternelle a chuté de près de 44 % au cours des 25 dernières années à l'échelle mondiale. De nombreux pays ont mis en place des politiques de santé qui ont contribué à cette évolution favorable. Ne pas prendre en compte la tendance naturelle des indicateurs de santé maternelle dans une étude d'impact d'une nouvelle intervention reviendrait à attribuer les changements observés uniquement à cette intervention et non à l'effet combiné du temps et de l'intervention. C'est pourquoi il est indispensable de comparer les résultats observés dans un groupe expérimental à ceux observés dans un groupe contrôle sans intervention extérieure. Nous verrons comment l'approche « différence des différences » permet dans ce cas d'individualiser l'effet de l'intervention de l'effet du temps.

Une autre implication méthodologique de la randomisation en grappe est en lien avec un déséquilibre fréquent des caractéristiques des individus entre les groupes de comparaison. En effet, la randomisation par grappe conduit de façon quasi systématique à une situation où la randomisation est antérieure à l'inclusion des sujets. Les caractéristiques des grappes sont alors réparties de façon homogène entre les groupes si la randomisation est effectuée dans les règles de l'art. Ce n'est pas nécessairement le cas en ce qui concerne les caractéristiques des individus qui constituent chaque grappe. Le déséquilibre qui en résulte, nommé ici « l'effet du groupe », peut expliquer les différences des indicateurs de santé entre les groupes de comparaison avant même d'avoir démarré l'intervention. Le risque de ne pas prendre en compte ce déséquilibre serait d'attribuer l'effet de l'intervention à la différence initiale entre les groupes, alors que l'intervention n'a par ailleurs aucune efficacité.

Enfin, en randomisant des grappes, on se met en situation de recueillir des données corrélées. Les sujets qui composent une unité « sociale » ou grappe ne peuvent être considérés comme indépendants les uns des autres. En effet les résultats observés chez deux sujets d'une même grappe auront tendance à être plus « similaires » que si les résultats étaient associés à deux sujets provenant de deux grappes distinctes (donc indépendantes). Si l'on ne tenait pas compte de la corrélation induite par le schéma expérimental (l'effet grappe), la puissance des tests statistiques se trouverait sous-estimée, les étendues des intervalles de confiance faussement diminuées et les conclusions de l'étude erronées.

Dans le cas de l'essai QUARITE réalisé au Sénégal et au Mali, nous avons randomisé des hôpitaux pour intervenir sur les professionnels de santé en faisant l'hypothèse que l'amélioration de la qualité des soins allait entrainer une diminution de la mortalité maternelle. Nous verrons, à travers cet exemple, comment l'utilisation des méthodes statistiques appropriées permet d'analyser correctement ces données.

Contexte et programme évalué : l'essai QUARITE pour réduire la mortalité maternelle hospitalière

Malgré la baisse de la mortalité maternelle dans le monde, celle-ci reste toujours très élevée dans de nombreux pays à faibles ou moyens revenus. Une large part des décès maternels est due à un traitement trop tardif ou inapproprié des complications obstétricales. Il est donc urgent de mettre en œuvre des interventions efficaces pour remédier à cette situation. Ces interventions consistent en général à mobiliser les communautés pour améliorer le recours aux soins en temps utile et à améliorer la filière de soins et les pratiques médicales aux différents niveaux de la pyramide sanitaire.

Nous avons testé par un ECR en grappes (essai QUARITE), entre 2007 et 2011, une approche combinant les revues de cas de mortalité maternelle dans les établissements de santé (ou audit de décès) et la formation continue du personnel de santé (Dumont *et al.* 2013). L'unité de randomisation et d'intervention était l'hôpital de référence dans lequel les complications obstétricales les plus sévères étaient prises en charge, au Mali comme au Sénégal. L'intervention ciblait les professionnel-le-s de santé qui travaillaient dans le même service (maternité). L'intervention d'une durée de deux ans

consistait à former initialement des leaders d'opinion locaux aux pratiques optimales et aux audits de décès maternels (formation de formateurs et formatrices). Les leaders d'opinion formaient ensuite leurs équipes (médecins, sages-femmes et infirmières qualifiées) dans chaque hôpital et implantaient les audits de décès maternels. Un facilitateur ou une facilitatrice externe visitait tous les trois mois chaque centre pour vérifier si l'agenda de formation était bien respecté et superviser une séance d'audit. L'audit des décès maternels avait pour objectif de rechercher de manière qualitative et approfondie les causes et les circonstances entourant les décès maternels survenus dans les établissements de santé. Dans un premier temps, ces décès étaient recensés et analysés dans les établissements. Mais lorsque cela était possible, ces études s'attachaient également à identifier les facteurs intervenant en dehors de l'hôpital, dans le système de santé ou dans la communauté, et ayant contribué au décès. La formation du personnel de santé ciblait les problèmes de prise en charge qui avait été identifiés lors des séances d'audit afin de contribuer à l'amélioration de la qualité des soins dans les hôpitaux participants.

Protocole d'évaluation

Pour mesurer les effets de l'intervention QUARITE sur la mortalité maternelle dans les 46 hôpitaux de référence participant à l'essai, nous avons enregistré toutes les naissances dans ces centres pendant quatre ans, entre octobre 2007 et octobre 2011. L'identification des morts maternelles a été réalisée à partir de multiples sources d'information comme les registres d'accouchement, du bloc opératoire, d'hospitalisation ou de la morgue et les dossiers cliniques disponibles. Le critère de jugement principal de l'essai était le taux de mortalité maternelle estimé par le nombre de décès maternels enregistré dans l'hôpital divisé par le nombre d'accouchements pendant la même période.

La collecte des données pendant quatre ans a permis d'estimer cet indicateur avant la mise en œuvre de l'intervention (année 1), pendant la période d'intervention (année 2 et 3) et après la fin de la l'intervention (année 4).

Les hôpitaux ont été randomisés à la fin de la première année de collecte, juste avant le démarrage de l'intervention, entre un groupe

expérimental exposé au programme et un groupe contrôle sans intervention extérieure. La figure 2 présente l'évolution de la mortalité maternelle dans les deux groupes pendant la période d'étude. Elle a diminué de 10,4 décès pour 1 000 accouchements à 6,8 décès pour 1000 dans le groupe expérimental entre les périodes « avant » et la période « après » l'intervention. Cette diminution a été moins marquée dans le groupe contrôle : entre 8,1 et 7,1 pour 1 000 accouchements.

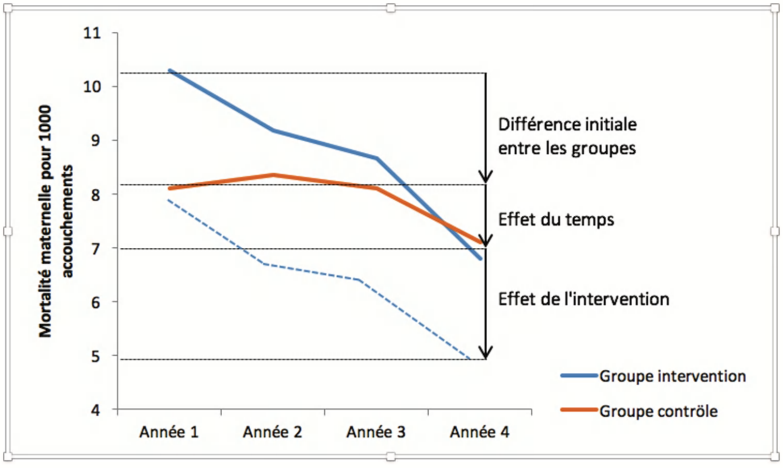


Figure 2. Évolution de la mortalité maternelle dans les deux groupes – essai QUARITE
Note : la ligne pointillée représente l'évolution de la mortalité maternelle dans le groupe d'intervention si son niveau de base (avant intervention) eut été équivalent à celui du groupe contrôle

Application de la méthode

Prise en compte de l'effet du temps : l'approche « différence des différences »

La différence de mortalité maternelle entre les deux périodes dans le groupe contrôle représente la tendance naturelle de la mortalité maternelle dans les hôpitaux de référence du Mali et du Sénégal (Figure 2 ci-dessus).

Cette évolution est peu marquée (-1,0 Décès pour 1 000 accouchements) sur la période d'étude, mais elle n'est pas négligeable. Elle témoigne de l'effet du changement des caractéristiques des patientes au cours du temps, mais aussi des changements intervenus dans le système de santé. La gratuité de la césarienne initiée en 2005 au Sénégal puis en 2006 au Mali a probablement facilité l'accès aux services de santé et amélioré progressivement l'issue des accouchements entre 2007 et 2011. De plus, la supervision régulière des équipes de soins pendant la période de l'essai pour renforcer la qualité des données collectées a peut-être contribué à améliorer les soins courants dans les hôpitaux des deux groupes et la prise en charge des patientes. Pour toutes ces raisons, les résultats de santé maternelle se sont améliorés au cours de l'étude dans le groupe contrôle alors que les hôpitaux de ce bras n'ont reçu aucune intervention de notre part. La méthode « différence des différences » (DD) permet de prendre en compte cet effet du temps dans l'étude d'impact.

Dans l'approche DD, l'effet de l'intervention est mesuré par la différence des changements observés entre les deux groupes (voir Tableau 3 ci-dessous). Les résultats peuvent être exprimés en termes de différence des risques de mortalité ou de rapport de cotes (RC).

En termes de différence des risques, l'effet de l'intervention correspond à la diminution des taux de mortalité maternelle dans le groupe expérimental (- 3,5 décès pour 1000) moins la diminution de mortalité dans le groupe contrôle (- 1,0 pour 1000). Dans cet exemple, la DD est donc estimée à (-3,5) – (-1,0) = – 2,5 décès pour 1 000 accouchements (IC à 95 % = – 3,5 ; – 1,5 pour 1 000 accouchements). En d'autres termes, l'intervention dans un hôpital réalisant 1 000 accouchements permettrait d'éviter 2,5 décès maternels par rapport à un autre hôpital de même volume d'activité, mais ne bénéficiant pas du programme. Cet effet est statistiquement significatif, car l'intervalle de confiance à 95 % de la DD ne contient pas 0.

Lorsque les résultats sont exprimés en rapport de cotes (RC), l'effet de l'intervention correspond au ratio entre les rapports de cotes des deux groupes. Dans notre cas, la cote pour un groupe donné à une période donnée est le nombre de décès maternels divisé par le nombre de femmes vivantes avant leur sortie de l'hôpital. Le rapport de cotes (RC) entre les deux périodes correspond donc à la variation du risque de mortalité maternelle entre l'année 1 et l'année 4. La mortalité ayant diminué dans les deux groupes entre les deux périodes d'étude, les RC sont inférieurs à 1 (Tableau 2). Toutefois,

cette diminution a été plus importante dans le groupe expérimental (RC = 0,66) que dans le groupe contrôle (RC = 0,89). Le rapport entre les deux RC, ici égal à 0,73 (IC à 95 % = 0,61 ; 0,91), mesure l'effet de l'intervention qui, là aussi, est statistiquement significatif, car l'intervalle de confiance du ratio des RC ne contient pas 1. En d'autres termes, l'intervention a permis de réduire la mortalité maternelle de 27 % par rapport aux changements observés dans le groupe contrôle.

***Tableau 3. Risques de mortalité maternelle selon les groupes et la période – essai* QUARITE**

	Année 1 Décès p. 1000 (effectifs)	Année 4 Décès p. 1000 (effectifs)	Différence de risques p. 1000 (IC à 95 %)	Rapport de cotes (IC à 95 %)
Groupe Intervention	10,4 (445/43269)	6,8 (356/52662)	– 3, 5 (- 2,3; – 4,7)	0,64 (0,57; 1,15)
Groupe Contrôle	8,1 (337/41655)	7,1 (381/53581)	– 1, 0 (- 0,1; – 2,0)	0,88 (0,76; 1,16)

Les résultats bruts estimés avec l'approche DD, que ce soit en termes de différence de risques ou de rapport de cotes, montrent que l'intervention est efficace pour réduire la mortalité maternelle hospitalière dans un système de santé tel que celui du Mali ou du Sénégal. Même si l'impact mesuré semble faible, ces résultats seraient suffisamment importants pour convaincre un décideur ou une décideuse de mettre en œuvre un tel programme à l'échelle de son pays sur l'ensemble des hôpitaux. Toutefois, cette personne pourrait remettre en cause les résultats de l'étude qui ne prennent pas en compte le déséquilibre initial entre les groupes ni le caractère corrélé des données. Nous verrons dans les deux paragraphes suivants comment prendre en compte l'effet du groupe et l'effet grappe.

Prise en compte de l'effet du groupe

La figure 2 ci-dessus montre que le niveau de la mortalité maternelle à l'année 1 était plus élevé dans les hôpitaux du groupe d'intervention que dans le groupe contrôle. La diminution plus rapide de la mortalité dans le bras expérimental a ramené ensuite le taux de décès maternel à un niveau quasi identique à celui du groupe contrôle à l'année 4. Si nous avions comparé les

résultats de l'essai entre les deux groupes à l'année 4, nous aurions conclu à tort que l'intervention n'avait aucun effet sur la mortalité.

L'approche DD, qui analyse simultanément les changements observés dans les deux groupes, a permis de pallier ce problème, mais elle n'a pas contrôlé le déséquilibre initial entre les groupes. Comme la randomisation des hôpitaux était antérieure à l'inclusion des patientes, leurs caractéristiques n'étaient pas réparties de façon homogène entre les bras de l'essai.

Tableau 4. Caractéristiques des patientes selon la période et le groupe – essai QUARITE

	Année 1		**Année 4**	
	Intervention (n=43269)	Contrôle (n=41 655)	Intervention (n=52662)	Contrôle (n=53581)
Réside loin de l'hôpital (en dehors de la région)	2 242 (5,2)	509 (1,2)	1 115 (2,1)	854 (1,6)
Évacuée d'un autre établissement	11 644 (26,9)	9 384 (22,5)	15 382 (29,2)	13 097 (24,4)
Âge ≥ 35 ans	4 356 (10,1)	4 195 (10,1)	5 325 (10,1)	5 210 (9,7)
Nullipare	28 435 (65,7)	26 939 (64,7)	35 908 (68,2)	36 397 (67,9)
Antécédents de césarienne	3 112 (7,2)	2 782 (6,7)	5 196 (9,9)	4 920 (9,2)
Pathologie avant la grossesse	406 (0,9)	324 (0,8)	516 (1,0)	648 (1,2)
Aucune visite prénatale	4 221 (9,8)	4 535 (10,9)	4 780 (9,1)	5 250 (9,8)
Pathologie pendant la grossesse	3 976 (9,2)	3 401 (8,2)	3 897 (7,4)	4 834 (9,0)
Grossesse multiple	1768 (4,1)	1555 (3,7)	2 139 (4,1)	2 135 (4,0)

Dans notre cas, la proportion des femmes qui résidaient loin de l'hôpital

où elles avaient accouché était plus importante dans le bras expérimental que dans le groupe contrôle, et de ce fait, les évacuations sanitaires étaient plus fréquentes parmi les hôpitaux bénéficiant de l'intervention (Tableau 4). Étant donné la faible performance des filières de soins au Mali comme au Sénégal et les retards accumulés dans la prise en charge lors des évacuations, l'état de santé des femmes du bras expérimental était globalement plus précaire et leur risque de mortalité plus élevé. Le déséquilibre entre les groupes semble s'atténuer à l'année 4. On peut alors se poser la question du biais que cela peut entrainer dans l'analyse d'efficacité de l'intervention selon l'approche DD : est-ce que la diminution plus importante de la mortalité maternelle dans le bras expérimental ne s'explique pas tout simplement par la modification du profil des patientes dans ce groupe ?

Plusieurs méthodes statistiques permettent de répondre à cette question en prenant en compte les facteurs de confusion dans l'étude d'impact. Les modèles de régression logistique sont particulièrement adaptés à ce type d'analyse, car ils permettent d'ajuster la mesure de l'effet (DD ou ratio des rapports de cotes) à des caractéristiques individuelles ou collectives (de grappe), telles que le lieu de résidence ou le mode d'admission des patientes dans l'hôpital. Nous verrons dans le paragraphe suivant que les modèles de régression logistique généralisés permettent de prendre en compte, dans le même temps, l'effet grappe.

Prise en compte de l'effet grappe

Puisque les patientes qui ont accouché dans le même hôpital ne sont pas totalement indépendantes les unes des autres, l'information rapportée par un hôpital de 1 000 femmes, par exemple, est donc moindre que s'il s'agissait de 1 000 patientes indépendantes ayant accouché dans différents centres. Il en résulte une perte de puissance dans l'analyse statistique qui doit être compensée par une augmentation du nombre de sujets à inclure. Par ailleurs, l'analyse des données recueillies doit recourir à des méthodes statistiques appropriées, sous peine d'obtenir des résultats abusivement significatifs. En effet, si l'on ne tenait pas compte de cet effet grappe, les étendues des intervalles de confiance s'en trouveraient faussement diminuées, car les tests statistiques ne prendraient pas en compte les variations inter- et intra-hospitalières de la mortalité maternelle. Différentes méthodes permettent

d'analyser correctement ce type de données parmi lesquelles l'utilisation de modèles mixtes ou marginaux. La commande *diff* du logiciel STATA (version 13.0), par exemple, permet d'estimer la DD en tenant compte de la structure hiérarchique des données, organisée en plusieurs niveaux imbriqués (niveau 1 : patientes ; niveau 2 : hôpital) et des éventuelles variables de confusions. La commande *xtlogit* du même logiciel permet d'estimer le ratio des rapports de cotes selon la même approche multivariée et multiniveaux.

Tableau 5. Estimations brutes et ajustées des effets de l'intervention – essai QUARITE

Méthode de mesure	**Estimation brute (IC à 95 %)**	**Estimation ajustée* (IC à 95 %)**
Différence de risques	-2,5 (- 3,5; – 1,5)	-2,5 (- 4,2; – 0,9)
Rapports de cotes	0,73 (0,61; 0,91)	0,85 (0,73; 0,98)

* effet ajusté sur l'effet grappe et sur les caractéristiques des patientes et des hôpitaux

Après ajustement sur les variables de confusion et sur l'effet grappe, le ratio des RC est passé de 0,73 (estimation brute) à 0,85 (estimation ajustée), comme le montre le tableau 5. En d'autres termes, le déséquilibre de certaines caractéristiques individuelles, telles que le niveau d'accessibilité à l'hôpital, explique en partie la diminution plus importante de la mortalité maternelle dans le bras expérimental comparé au bras témoin, mais pas totalement. On peut donc attribuer une diminution relative de 15 % de la mortalité maternelle à l'intervention. De plus, après ajustement sur l'effet grappe, les étendues des intervalles de confiance sont plus importantes que dans les estimations brutes, quelle que soit la méthode de mesure utilisée (différence de risques ou rapport de cotes). On a donc perdu un peu de précision dans les mesures d'impact, mais les différences restent significatives compte tenu de la taille importante des échantillons.

Analyse réflexive et conclusion

Nous avons vu à travers ce chapitre l'intérêt de mener un ECR en grappe pour évaluer l'impact des interventions en santé maternelle et comment

contrôler les trois principaux biais dans la mesure d'efficacité : l'effet du temps, l'effet du groupe et l'effet grappe. Il existe d'autres particularités des essais en grappe que nous aborderons dans la discussion sans toutefois les détailler. Ces particularités tiennent aux questions suivantes : l'expérimentation se fait-elle en aveugle ? Le consentement des sujets est-il nécessaire ? Les résultats sont-ils généralisables à d'autres contextes ?

Lorsque l'intervention que nous voulons tester prend la forme d'un programme d'activités ciblant des grappes d'individus (comme le programme d'amélioration de la qualité de soins dans l'essai QUARITE qui ciblait les professionnel-le-s de santé), l'inclusion des sujets se fait alors inévitablement en ayant connaissance du bras dans lequel il sera inclus. L'essai est dit « ouvert », contrairement à la plupart des essais individuels réalisés en « aveugle », c'est-à-dire sans avoir connaissance du bras dans lequel les sujets sont répartis aléatoirement. Cette situation est potentiellement source de biais qui peuvent être induits par des comportements éventuellement différents entre les groupes.

La question du recueil de consentement pour participer à l'étude est également importante. Si le consentement du ou de la responsable de grappe est incontournable, celui des sujets n'est pas nécessairement requis. Il est possible d'informer les individus susceptibles de participer à l'étude, sans toutefois obtenir un consentement signé de leur part. Dans l'essai QUARITE, par exemple, nous avons obtenu le consentement du directeur d'hôpital et du chef de service de la maternité. En revanche, nous ne l'avons pas fait pour les femmes ayant accouché dans chaque hôpital participant à l'étude, puisque nous n'intervenions pas directement sur elles et que les données collectées anonymisées étaient issues des registres hospitaliers existants, sans interview auprès des femmes.

Une autre singularité de l'essai en grappe tient au fait que le contexte dans lequel l'intervention est mise en œuvre peut varier considérablement d'une grappe à l'autre. Si l'on connaît à l'avance les caractéristiques des grappes qui ont un effet sur le critère de jugement principal, il est d'ailleurs recommandé de stratifier la randomisation sur ces critères. Dans le cas de l'essai QUARITE, nous avons randomisé des hôpitaux en fonction du niveau de soins, car les données disponibles préliminaires nous indiquaient que les taux de mortalité étaient différents entre les hôpitaux de district, les hôpitaux régionaux et les hôpitaux de la capitale. L'analyse des résultats selon le niveau de soins montrait que l'intervention était particulièrement

efficace dans les hôpitaux de district et peu efficace dans les hôpitaux régionaux. Nous avons donc conclu que les bénéfices de l'intervention devaient s'appliquer uniquement aux hôpitaux de première ligne et que d'autres recherches seraient nécessaires pour confirmer ou non l'utilité d'un tel programme dans des hôpitaux de niveau supérieur.

Même s'ils sont complexes à mettre en œuvre, les essais contrôlés randomisés sont les études d'impact en santé maternelle et périnatale qui fournissent le niveau de preuve le plus élevé s'ils sont menés dans les règles de l'art. Pour des raisons logistiques, organisationnelles et parfois éthiques, il est souvent plus pertinent de randomiser des unités « sociales » ou grappes que des individus. Ces essais en grappes entrainent des défis méthodologiques qui ne sont pas insurmontables à condition de les avoir pris en compte dans la mise en place de l'étude et dans les analyses. On doit également garder à l'esprit que, dans ce type d'étude, l'unité de randomisation est différente de l'unité l'analyse, c'est-à-dire l'unité qui sera prise en compte pour l'interprétation des résultats ou analyse d'impact. Mais le niveau de preuve d'efficacité d'une intervention n'est pas l'unique critère de choix pour les décideuses et décideurs. Le contexte et les conditions de mise en œuvre sont aussi des éléments essentiels pour concevoir des politiques publiques plus efficaces. Les futures recherches devraient permettre également de comprendre comment et pourquoi cette intervention produit des effets positifs sur la santé des mères et des nouveau-nés. Les essais hybrides combinant à la fois les essais contrôlés en grappe et l'analyse d'implantation sont les plus adaptés pour répondre simultanément aux questions d'efficacité et de mise en œuvre.

Références clés

Campbell, M. K., Mollison, J., Steen, N., Grimshaw, J. M. et Eccles, M. (2000) Analysis of cluster randomized trials in primary care: a practical approach. *Family Practice*, 17(2), 192–196.

Un article qui présente très clairement la méthode des essais randomisés en grappe.

Dumont, A., Fournier, P., Abrahamowicz, M., Traoré, M., Haddad, S. et Fraser, W. (2013). Quality of care, risk management, and technology in obstetrics : a cluster-randomized trial of a multifaceted intervention to reduce

hospital-based maternal mortality in Senegal and Mali (QUARITE). *Lancet*, 382(9887), 146 – 157.

Les résultats de cette étude réalisée au Sénégal et au Mali ont été utilisés pour illustrer ce chapitre.

Chaillet, N., Dumont, A., Abrahamowicz, M., Pasquier, J. C., Audibert, F., Monnier, P., ... et Fraser, W. D. (2015). A cluster-randomized trial to reduce cesarean delivery rates in Quebec. *New England Journal of Medicine*, 372(18), 1710-1721.

Les résultats de cette étude réalisée au Canada permettent d'illustrer les défis méthodologiques liés aux essais en grappe.

Références

Dumont, A., Fournier, P., Abrahamowicz, M., Traoré, M., Haddad, S. et Fraser W. (2013). Quality of care, risk management, and technology in obstetrics : a cluster-randomized trial of a multifaceted intervention to reduce hospital-based maternal mortality in Senegal and Mali (QUARITE). *Lancet*, 382(9887), 146–157

Stanton, C. K., Newton, S., Mullany, L. C., Cofie, P., Agyemang C. T., Adiibokah, E., ... et Gyapong, J. (2013). Effect on postpartum hemorrhage of prophylactic oxytocin (10 IU) by injection by community health officers in Ghana: A community-based, cluster-randomized trial. *PLoS Medicine*, 10(10), e1001524.
doi: 10.1371/journal.pmed.1001524

Villar, J., Ba'aqeel, H., Piaggio, G., Lumbiganon, P., Belizán, J. M, Farnot, U., ... et Garcia, J. (2001). WHO antenatal care trial research group. *Lancet*, 357(9268), 1551-64.

Résumé / Abstract / Resumen

Les essais contrôlés randomisés en grappe permettent d'évaluer l'impact des interventions en santé avec un niveau de preuve très élevé. Si les essais randomisés en grappes offrent plusieurs avantages dans les pays à faibles ressources par rapport aux essais individuels, ils présentent quelques défis

méthodologiques qui tiennent au fait que l'unité d'analyse est différente de l'unité de randomisation ou d'intervention. Ces défis tant sur le plan de la mise en œuvre de l'intervention que de l'analyse statistique sont présentés à partir d'une expérience réalisée au Mali et au Sénégal (essai QUARITE) pour réduire la mortalité maternelle et néonatale hospitalière. Nous verrons comment l'approche « différence des différences » permet de mesurer l'effet l'intervention tout en tenant compte de l'effet du temps qui joue souvent favorablement sur les indicateurs de santé en dehors de toute intervention extérieure. Les modèles statistiques appropriés sont présentés pour pouvoir ajuster l'effet de l'intervention sur les différences qui existent entre les groupes expérimentaux avant le démarrage du programme à tester et sur l'effet de groupe propre aux essais en grappe.

***

Cluster-randomised trials are used to assess the impact of health interventions with a very high level of evidenceexpected. While cluster randomized trials offer several advantages in low-resource countries compared to individual trials, they present some methodological issues because the unit of analysis is different from the unit of randomization and intervention. These challenges, both in terms of the intervention implementation and the statistical analysis, are presented based on a trial conducted in Mali and Senegal (QUARITE trial) to reduce maternal and neonatal mortality in hospitals. We will see how the difference-in-differences approach measures the effect of the intervention taking into account the secular trends, which often plays favorably on health indicators outside of any external intervention. The appropriate statistical models are presented in order to adjust the effect of the intervention on baseline case-mix and on the cluster effect which is specific to this type of trial.

***

Los ensayos controlados aleatorios grupales permiten evaluar el impacto de las intervenciones de salud con un nivel muy alto de pruebas. Aunque los ensayos aleatorios grupales ofrecen varias ventajas en los países de bajos recursos con respecto a los ensayos individuales, presentan algunos desafíos metodológicos debido al hecho de que la unidad de análisis es diferente de la unidad de aleatorización o de intervención. Estos desafíos, tanto en la implementación de la intervención como en el análisis estadístico, se presentan sobre la base de un experimento realizado en Malí y Senegal

(ensayo QUARITE) para reducir la mortalidad materna y neonatal en los hospitales. Veremos cómo el enfoque de « diferencia de diferencias » permite medir el efecto de la intervención teniendo en cuenta el efecto del tiempo, que a menudo tiene un impacto positivo en los indicadores de salud fuera de cualquier intervención externa. Se presentan modelos estadísticos apropiados para ajustar el efecto de la intervención sobre las diferencias entre los grupos experimentales antes del inicio del programa de prueba y sobre el efecto del grupo específico de los ensayos de grupos.

***

Alexandre Dumont est gynécologue-obstétricien et épidémiologiste. Il a mené des essais contrôlés en grappe en santé maternelle et périnatale au Canada, au Sénégal, au Mali et au Burkina Faso. Il a reçu le prix « Birthday Prize » de la Commission européenne pour l'essai QUARITE qui a permis de réduire la mortalité maternelle et néonatale dans les hôpitaux de référence du Mali et du Sénégal.

Citation

Alexandre Dumont. (2019). L'évaluation de la pérennité. Un exemple en santé maternelle et infantile. In *Évaluation des interventions de santé mondiale. Méthodes avancées*. Sous la direction de Valéry Ridde et Christian Dagenais, pp. 263-280. Québec : Éditions science et bien commun et Marseille : IRD Éditions.

11. La mesure de l'équité

Un exemple d'intervention de gratuité des soins obstétricaux

TARIK BENMARHNIA ET BRITT MCKINNON

La mesure de l'équité

Définition de la méthode

Nous définissons l'équité des effets d'une intervention en fonction de l'identification d'une hétérogénéité dans les bénéfices potentiels entre différents sous-groupes ou territoires d'une population. Une intervention sera évaluée comme équitable si les bénéfices sont plus importants au sein des sous-groupes ou territoires qui auront été définis comme les plus vulnérables ou ayant les besoins les plus importants.

Forces de la méthode

La méthode proposée permet de quantifier les inégalités de santé pour plusieurs configurations (groupes ordonnés ou non) sur une échelle absolue ou relative.

Défis de la méthode

Il est important de disposer d'un bon niveau de connaissance des populations vulnérables et de bien les étudier empiriquement et théoriquement. Pour certaines des méthodes proposées (par exemple, les analyses de décomposition...), il est parfois nécessaire d'avoir de grands échantillons pour identifier une hétérogénéité dans les bénéfices potentiels d'une intervention au sein de deux ou plusieurs groupes.

L'équité en santé est largement identifiée comme un objectif auquel les politiques et activités de santé publique doivent répondre en priorité (Starfield 2007; Marmot, Friel *et al.* 2008; Blas et Kurup 2010). Toutefois, il existe une grande diversité dans les définitions de la notion d'équité en santé et de telles implications sont amplement discutées (Whitehead 1992; Braveman et Gruskin 2003; Braveman, Cubbin *et al.* 2005). Parfois, cette notion, souvent définie via son opposé qui est l'inéquité ou iniquité, se réfère à la description de différences évitables ou remédiables entre différents groupes de personnes (Whitehead 1992; Kawachi, Subramanian *et al.* 2002), incluant donc une dimension morale et éthique (Daniels 2006; Smith 2015). Parfois, elle fait référence aux conséquences d'une intervention de santé publique. En effet, dans cette dernière dimension, la notion d'inéquité implique une intervention qui viendra accroitre des inégalités de santé déjà existantes. Ces deux notions sont parfois fusionnées, mais souvent les

implications et hypothèses sur le plan des théories de justice sociale sous-jacentes sont trop peu discutées (Smith 2015).

Dans ce chapitre, nous proposons une définition de l'équité en santé qui fait référence à la distribution des bénéfices des politiques publiques.

Cette définition implique, dans un premier temps, l'identification des sous-groupes ou territoires d'une population pour lesquels l'état de santé sera considéré comme plus négatif ou au sein desquels l'impact d'un déterminant de la santé sera plus important (notion épidémiologique de modification d'effet) en comparaison à un autre sous-groupe ou territoire de la même population. Nous appellerons ces sous-groupes d'une population « populations vulnérables », en comparaison avec des populations non vulnérables.

Ainsi, dans ce chapitre, nous définissons l'équité des effets d'une intervention en fonction de l'identification d'une hétérogénéité dans les bénéfices potentiels entre différents sous-groupes ou territoires d'une population. Une intervention sera évaluée comme équitable si les bénéfices sont plus importants au sein des sous-groupes ou territoires qui auront été définis comme vulnérables.

La notion d'équité en santé est discutée depuis plusieurs décennies, mais relativement peu mise en œuvre. Entendue dans le contexte d'une distribution inégale des effets d'interventions de santé publique, son évaluation est minoritaire en comparaison avec l'évaluation des effets sur toute une population. Pour illustrer cela, nous pouvons prendre deux exemples. Les évaluations des effets des politiques publiques de lutte contre le tabagisme sont massivement étudiées dans la littérature épidémiologique (Hoffman et Tan 2015) et se comptent par centaines. Cependant, la grande majorité de ces études portent principalement sur une population dans son ensemble (Frazer, Callinan *et al.* 2016) ou sur un sous-groupe de la population comme les adolescent-e-s (Gilman, Rende *et al.* 2009) ou certains professionnel-le-s (Farrelly, Nonnemaker *et al.* 2005). Les études qui évaluent les inégalités des effets d'une politique publique de lutte contre le tabagisme entre deux ou plusieurs sous-groupes d'une population sont bien plus rares (Brown, Platt *et al.* 2014a; Brown, Platt *et al.* 2014b) et concernent par exemple les inégalités selon le sexe, l'âge ou le niveau d'éducation. Le deuxième exemple porte sur les interventions de lutte contre la pollution atmosphérique en milieu urbain (Cartier, Benmarhnia *et al.* 2015). Dans une

revue systématique (Benmarhnia, Rey *et al.* 2014), nous avons pu mettre en évidence que sur plus de 50 études publiées, seulement deux études ont formellement évalué l'hétérogénéité des effets de l'intervention au sein de différents sous-groupes et huit études permettaient une évaluation *a posteriori* entre les estimations stratifiées. Ces deux exemples mettent en évidence, d'une part, le faible nombre d'études ayant investigué cette question de recherche spécifique, malgré la grande disponibilité des données permettant de mener de telles études, et d'autre part, le manque de méthodes appropriées permettant d'évaluer l'équité en santé, la plupart des études basant leurs conclusions sur des comparaisons qualitatives entre différentes strates de population ou via un test d'interaction dans un modèle de régression.

Dans ce chapitre nous présentons différentes méthodes de mesures des inégalités de santé et leur application à l'évaluation des effets des interventions à travers l'exemple de l'effet de l'exemption du paiement des frais de soins sur l'accès aux soins obstétricaux dans le cadre d'un exemple simulé.

Nous présentons dans un premier temps le contexte d'évaluation en décrivant l'intervention considérée, les données étudiées et les justifications d'un potentiel effet hétérogène de cette intervention sur l'accès aux soins obstétricaux. Ensuite, nous présentons différentes méthodes pour mesurer les inégalités de santé et différentes approches pour les appliquer à l'évaluation des effets des interventions, donc à l'équité telle que définie dans ce chapitre.

Présentation du contexte d'évaluation de l'équité : l'exemple de l'exemption du paiement des frais de soins (gratuité des soins) et l'impact sur l'accès aux soins obstétricaux

L'accès aux soins de base demeure un enjeu fondamental dans de nombreux pays à faible et moyen revenus (Althabe, Bergel *et al.* 2008; Lagarde et Palmer 2008; Dzakpasu, Powell-Jackson *et al.* 2013; McKinnon, Harper *et al.* 2014; McKinnon, Harper *et al.* 2015). Dans le contexte de systèmes de santé largement sous-financés, ces pays se retrouvent face à un dilemme. D'une part, le paiement des frais de soins peut présenter une barrière financière à l'accès aux soins. D'autre part, une diminution des

ressources financières des systèmes de santé peut contribuer à une baisse de la qualité des services de santé qui pourra aussi constituer des barrières financières à l'assistance qualifiée à l'accouchement. Certaines études ont montré que la mise en place du paiement des frais de soins contribue à augmenter les ressources au niveau local et à améliorer la qualité des services (Ellis 1987; Litvack et Bodart 1993). D'autres études ont montré des effets négatifs, notamment sur le plan de l'équité (Yoder 1989; Russell et Gilson 1997). De nombreuses campagnes internationales visant à promouvoir la gratuité des soins (par exemple l'exemption de la tarification des accouchements) ont été menées (Yates 2009; Ridde et Morestin 2011).

Dans ce contexte, nous proposons de prendre comme exemple d'intervention la mise en place de la gratuité des soins obstétricaux (c'est-à-dire l'exemption de la tarification des accouchements) et d'évaluer son impact et son équité quant à l'accès aux soins obstétricaux. Afin d'évaluer les effets vis-à-vis de l'équité, nous allons considérer deux dimensions quant à une potentielle hétérogénéité dans les bénéfices de cette intervention. Premièrement, nous allons considérer une dimension liée aux inégalités matérielles que nous allons mesurer par un indice de richesse (*wealth index*) souvent utilisé dans les enquêtes populationnelles internationales (Filmer et Pritchett 1999; Howe, Hargreaves *et al.* 2009). Deuxièmement, nous allons considérer une dimension liée aux inégalités sociales par la mesure du niveau d'éducation de la mère. Ce choix est justifié par les inégalités sociales et matérielles quant à l'accès aux soins obstétricaux dans un contexte de pays à faible et moyen revenus (Say et Raine 2007). Ainsi, nous définissons l'intervention comme équitable si elle contribue à réduire les inégalités matérielles et/ou sociales d'accès aux soins obstétricaux.

L'exemple que nous présentons ci-dessous provient de données simulées à partir d'enquêtes populationnelles et ne correspond pas à une situation réelle d'intervention. De nombreux exemples de ce type d'intervention existent (Lagarde et Palmer 2008). Nous nous concentrons sur les inégalités socio-économiques, mais la définition d'équité adoptée dans ce chapitre peut s'appliquer également à d'autres facteurs démographiques (âge, sexe, genre) ou territoriaux (urbain vs rural).

Méthodes pour mesurer les inégalités de santé

Un certain nombre d'indicateurs peuvent être utilisés pour mesurer les inégalités de santé. Toutefois, il n'existe pas de consensus quant à une mesure en particulier, notamment lorsqu'il s'agit de mesurer les variations d'inégalités dans le temps. Cette section commence par une discussion sur des enjeux importants quant à la mesure des inégalités, notamment sociales, en santé. Dans la section suivante, nous allons décrire différents indicateurs permettant de mesurer des inégalités de santé avec des exemples les illustrant.

Enjeux quant à la mesure des inégalités de santé

Inégalités absolues vs relatives

Il est important tout d'abord de bien distinguer les mesures d'inégalités relatives et absolues dans la mesure où elles fournissent toutes deux des informations importantes, mais fondamentalement différentes vis-à-vis de la présence et de l'ampleur d'inégalités. Les mesures absolues reflètent l'ampleur des différences de santé entre deux groupes. Par exemple, les taux de mortalité pour cause cardiovasculaire en 2014 étaient de 12 pour 1000 hommes et de 8 pour 1000 femmes, la mesure absolue d'inégalité serait de 4 pour 1000 x année. En parallèle, les mesures relatives reflètent une différence proportionnelle de santé entre deux groupes. Dans l'exemple précèdent, elle représenterait le ratio des deux taux, à savoir 12/8 = 1,5. Parce que ces deux mesures peuvent conduire à des conclusions distinctes vis-à-vis de l'ampleur et des changements concernant les inégalités, il est important de rapporter les deux échelles pour présenter un portrait complet des inégalités (Lynch, Smith *et al.* 2006; Houweling, Kunst *et al.* 2007; Harper, King *et al.* 2010).

Classifier les groupes sociaux

Certains groupements sociaux ont « naturellement » deux groupes

(homme/femme, urbain/rural), tandis que d'autres groupes possèdent plusieurs niveaux (niveau d'éducation, ethnicité). Pour les groupements sociaux à plusieurs groupes, le choix des indicateurs d'inégalités à adopter se fera en fonction du fait que les groupes sont organisés selon une structure ordonnée ou non. Le niveau d'éducation ou le niveau d'indice de richesse constituent par exemple des groupements ordonnés, tandis que l'ethnicité ou des caractéristiques géographiques ne sont pas *a priori* ordonnées. Des comparaisons par paire peuvent systématiquement être menées entre deux groupes pour n'importe quel type de variable sociale. De plus, l'ampleur des inégalités sociales pour plusieurs groupes (ordonnés ou non) peut être mesurée à partir d'indicateurs plus complexes comme l'indice d'inclinaison des inégalités (*slope index of inequalities*) ou l'indice de Theil.

Mesures des inégalités de santé

Les indicateurs simples d'inégalités (comparaison par paires)

Les mesures standards pour comparer l'état de santé entre deux groupes sociaux sont la différence et le ratio. Ces deux mesures sont complémentaires. Si l'indicateur de santé est exprimé en taux (par exemple la mortalité par cancer pour 100 000 personnes par an), les mesures absolues et relatives correspondent respectivement à une différence de taux et un ratio de taux, comparant deux groupes sociaux d'intérêt.

Pour les groupes sociaux ordonnés via le niveau d'éducation par exemple, les comparaisons par paires se font en général par la comparaison des groupes aux extrémités. Par exemple, si l'on considère un indice de richesse stratifié en quintiles, il s'agira de comparer le quintile le plus riche au quintile le plus pauvre. Pourtant, ce type de comparaison ignore l'état de santé des trois autres quintiles, qui peut fournir une information importante vis-à-vis d'inégalités au sein de la population. La figure 1 présente la proportion de femmes ayant reçu des soins obstétricaux auprès d'un-e professionnel-le qualifié-e par quintile d'un indice de richesse pour deux populations hypothétiques. Les deux populations ont les mêmes valeurs de différence ou de ratio en comparant les deux quintiles extrêmes, tandis que le portrait au sein des trois autres quintiles est différent. C'est la raison pour

laquelle des mesures d'inégalités plus complexes, considérant l'ensemble des groupes, sont généralement préférées.

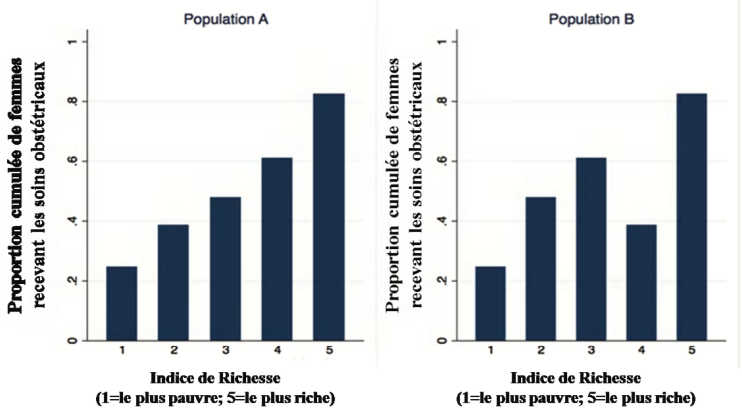


Figure 1. Proportion de femmes ayant reçu des soins obstétricaux auprès d'un professionnel qualifié par quintile d'un indice de richesse pour deux populations hypothétiques

L'indice d'inclinaison des inégalités et l'indice de concentration sont deux mesures souvent utilisées pour mesurer le gradient de santé au sein de groupes sociaux ordonnés. Contrairement aux mesures simples présentées ci-dessus, ces indices prennent en compte la taille relative de chaque groupe.

L'indice d'inclinaison des inégalités

L'indice d'inclinaison des inégalités mesure la différence absolue par rapport à des valeurs prédites entre les groupes les plus hauts et les groupes les plus bas de la distribution de la position sociale pour des groupes ordonnés (éducation, indice de richesse, etc.) (Pamuk 1985). Cet indice est estimé à partir du modèle de régression approprié à la mesure de santé considérée (par exemple, une régression linéaire pour une mesure continue, et une régression logistique pour une mesure de santé binaire). Il est possible de calculer cet indice en utilisant l'indice de richesse en continu par exemple, en allant des plus pauvres (rang=0) aux plus riches (rang=1). La mesure de

santé est ensuite incluse dans un modèle de régression en incluant, comme prédicteur, le rang de chaque individu au sein de la distribution cumulative de richesse. Ainsi, le coefficient correspondant à ce rang représentera la différence de santé estimée entre les plus pauvres et les plus riches individus de la population.

La figure 2 est une représentation graphique de cet indice, montrant la valeur prédite de l'issue de santé (ici le fait de recevoir des soins obstétricaux par un-e professionnel-le qualifié-e) en fonction de la proportion cumulée du rang par rapport au niveau d'éducation. La valeur prédite de prévalence de réception de soins obstétricaux (par un-e professionnel-le qualifié-e) parmi les femmes avec le plus bas niveau d'éducation (rang=0) est de 17,7%. Parmi les femmes avec le plus haut niveau (rang=1), cette valeur est de 84%. Ainsi, la valeur de cet indice sera de 84-17.7 = 66,3 en %, ce qui signifie que les femmes au plus haut de la distribution du niveau d'éducation ont 63% plus de chance de recevoir des soins obstétricaux par un-e professionnel-le qualifié-e en comparaison avec les femmes au plus bas de cette distribution.

Lorsque les individus sont classés des plus défavorisés au moins défavorisés, des valeurs positives de cet indice indiquent que la prévalence de la mesure de santé est plus importante au sein des individus les moins défavorisés alors qu'une valeur négative indique que cette prévalence est plus importante au sein des plus défavorisés. Une valeur de 0 signifie une égalité au sein du gradient social. Une mesure relative de cet indice peut également être estimée en calculant le ratio entre les valeurs prédites des deux extrémités de la distribution. Selon la figure 2, cet indice sur l'échelle relative serait de 84/17,7=4,7, signifiant que la réception de soins obstétricaux par un-e professionnel-le qualifié-e est 4,7 fois plus élevée au sein du groupe le plus favorisé en comparaison au groupe le moins favorisé.

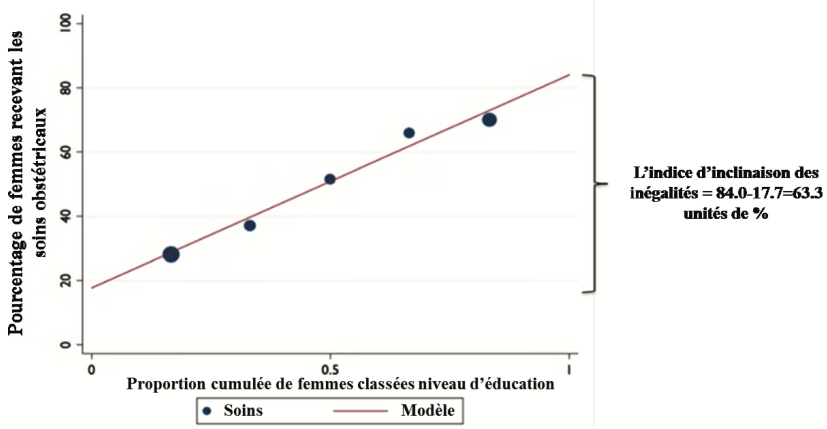


Figure 2. L'indice d'inclinaison des inégalités montrant les inégalités vis-à-vis des soins obstétricaux en fonction du niveau d'éducation.

L'indice de concentration

L'indice de concentration est une mesure relative des inégalités basée sur la courbe de concentration, une figure représentant la proportion cumulée d'individus classés par rang de santé sur l'axe X avec la proportion cumulée sur l'axe Y. La figure 3 représente la courbe de concentration avec les inégalités de richesse vis-à-vis de soins obstétriques utilisant les données présentées dans le tableau 1. La ligne à 45° représente la ligne d'égalité lorsque les ressources sont distribuées également au sein de la population. La courbe de concentration est située en dessous de cette ligne d'égalité, indiquant une inégalité défavorable aux femmes les plus pauvres. Par exemple, on peut observer que les 40% des femmes les plus pauvres ne bénéficient que de 25% à peu près du total des visites obstétricales.

Tableau 1. Calcul des valeurs de proportion cumulée pour générer la courbe de concentration

Indice de Richesse	Nombre de femmes	Proportion de femmes	Proportion cumulée de femmes	Nombre de femmes recevant les soins	Proportion de femmes recevant les soins	Proportion cumulée de femmes recevant les soins
1	1167	0,217	0,217	288	0,105	0,105
2	1017	0,190	0,407	394	0,143	0,248
3	990	0,184	0,592	475	0,172	0,420
4	989	0,184	0,776	604	0,219	0,640
5	1203	0,224	1,0	992	0,360	1,0

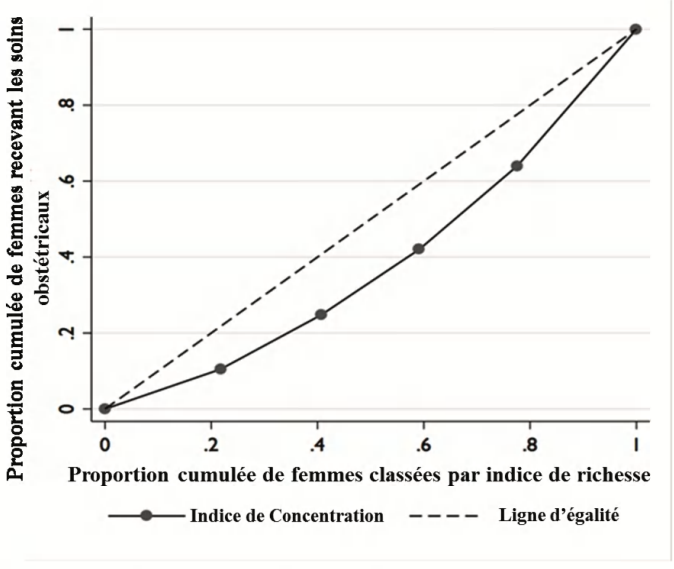


Figure 3. Courbe de concentration représentant les inégalités relatives de richesse vis-à-vis de soins obstétricaux

L'indice de concentration est défini comme le double de l'aire située entre la courbe de concentration et la ligne d'égalité (Wagstaff, Van Doorslaer *et al.* 1991). Cet indice peut se situer entre -1 et 1, avec 0 représentant une égalité complète. Des valeurs positives signifient des

inégalités au détriment des plus pauvres pour une mesure de santé positive (comme dans cet exemple). Pour des données regroupées par quintile par exemple, comme dans le tableau 2, l'indice de concentration se calcule grâce à la formule suivante :

$$C = (p_1 L_2 - p_2 L_1) + (p_2 L_3 - p_3 L_2) + \ldots + (p_{T-1} L_T - p_T L_{T-1})$$

où p_t est le pourcentage cumulé de l'échantillon classé par indice de richesse pour le groupe t, et L_t correspond la valeur de répartition de la mesure de santé par quintile (O'Donnell et Wagstaff 2008). Le tableau2 utilise les données de la courbe de concentration présentée ci-dessus et la dernière colonne présente le calcul de l'indice de concentration. Ici cet indice est de 0,224.

Tableau 2. Calcul de l'indice de concentration

Indice de Richesse (t)	**Proportion cumulée de femmes par quintile de richesse (pt)**	**Proportion cumulée de femmes recevant les soins (Lt)**	**Indice de Concentration**
1	0,217	0,105	0,011
2	0,407	0,248	0,025
3	0,592	0,420	0,052
4	0,776	0,640	0,136
5	1,0	1,0	0,000
			0,224

L'indice de concentration est une mesure d'inégalité relative, mais il est possible d'obtenir une version absolue de cet indice en multipliant l'indice de concentration obtenu par la moyenne de la mesure de santé. Il est aussi possible de calculer cet indice de concentration au niveau individuel en utilisant des approches basées sur des régressions[1]

Un enjeu important dans la recherche portant sur les inégalités de

1. Pour plus d'information, voir Kakwani, Wagstaff *et al.* (1997) et O'Donnell et Wagstaff (2008).

santé concerne la compréhension des facteurs qui contribuent le plus aux inégalités par rapport à une mesure de santé donnée. Un avantage de cet indice est qu'il peut être utilisé dans le cadre d'analyses de décomposition. Ce type d'analyses quantifie la contribution de différents facteurs quant à des inégalités sociales vis-à-vis d'une mesure de santé. Plusieurs exemples existent dans la littérature allant de la compréhension des déterminants des inégalités sociales de l'obésité au Canada (Hajizadeh, Campbell *et al.* 2014) à celle des inégalités sociales de mortalité infantile en Iran (Hosseinpoor, Van Doorslaer *et al.* 2006)[2].

Mesures complexes des inégalités avec groupes non ordonnés – l'indice de Theil

Les deux indices présentés ci-dessus sont valables uniquement lorsque les groupes sociaux sont ordonnés. Lorsque les groupes sociaux ne sont pas ordonnés, il est possible d'utiliser l'indice de Theil (Theil 1967). Les mesures d'inégalités absolues pour des catégories non-ordonnées incluent la méthode de variance (Harper, Lynch *et al.* 2008) entre groupes et l'indice de disparité (Pearcy et Keppel 2002) récemment étendue aux analyses de séries temporelles (Benmarhnia, Grenier *et al.* 2015).

L'indice de Theil est estimé à partir de la formule $T = \sum_{i=1}^{n} p_i \; r_i \ln (r_i)$ où p_i est la proportion de la population dans le groupe social i et r_i est le ratio de la moyenne de la mesure de santé dans le groupe i par rapport à la moyenne pour l'ensemble des groupes. Ainsi, si chaque groupe a une valeur moyenne de la mesure de santé égale à la moyenne pour tous les groupes, alors il n'y a pas d'inégalité relative et l'indice de Theil sera égal à 0. Cet indice est toujours positif. Plus sa valeur est grande, plus le niveau d'inégalité relative entre les groupes est important.

Les données dans le tableau 3 montrent comment calculer l'indice de Theil pour mesurer une inégalité relative d'accès aux soins obstétricaux au sein de cinq régions géographiques distinctes.

2. Pour plus d'informations sur les analyses de décomposition de l'indice de concentration, voir Wagstaff, Van Doorslaer *et al.* (1991), Wagstaff, Van Doorslaer *et al.* 2003 et O'Donnell and Wagstaff (2008).

Tableau 3. Calcul de l'indice de Theil d'accès aux soins obstétricaux au sein de cinq régions géographiques

Région	Soins obstétricaux (%)	Proposition de population (p_i)	Ratio de % de soins dans la région par rapport à la population totale (r_i)	$\ln(r_i)$	Indice de Theil
A	45,8	0,29	0,90	-0,11	-0,028
B	55,7	0,12	1,09	0,09	0,011
C	48	0,18	0,94	-0,06	-0,010
D	82,5	0,22	1,62	0,48	0,175
E	22,8	0,19	0,45	-0,81	-0,067
Toutes régions	51,0				**0,081**

Dans cette section, nous avons présenté un portrait de différentes mesures d'inégalités de santé. Il ne s'agit ni d'une liste exhaustive des mesures existantes pour mesurer les inégalités de santé, ni d'une discussion approfondie des enjeux conceptuels. Pour ce dernier point, nous recommandons les travaux de Kawachi et Kennedy (1997), Mackenbach et Kunst (1997) et Regidor (2004). Toutefois, nous avons pu présenter plusieurs cas de figure qui pourront ensuite être utilisés pour évaluer la distribution des effets d'interventions de santé publique tels que décrits dans la section suivante.

Application aux études d'évaluation des effets

Dans cette partie, nous allons présenter comment appliquer les méthodes de mesures des inégalités présentées ci-dessus à l'évaluation des effets d'une intervention de santé publique.

Nous présentons deux types d'approches : une approche d'évaluation simple permettant d'illustrer facilement les méthodes présentées ci-dessus et qui nous permettra également d'introduire les tests d'hétérogénéité et une approche basée sur les designs quasi-expérimentaux. Les détails des méthodes d'évaluation, particulièrement quant aux approches quasi-expérimentales, sont présentés dans le chapitre précédent.

Approches simples de type avant/après

Une première approche simple permettant d'évaluer si les bénéfices d'une intervention (ici la mise en place de la gratuité des soins obstétricaux) sont équitables est de comparer si les inégalités d'accès aux soins obstétricaux entre différents groupes sont modifiées avant et après la mise en place de l'intervention. Nous pouvons prendre comme exemple la comparaison du premier quintile de l'indice de richesse au cinquième quintile de cet indice. Ainsi en calculant la différence ou le ratio entre ces deux groupes, d'une part, avant la mise en place de l'intervention et, d'autre part, après sa mise en place, nous pouvons obtenir deux mesures (sur l'échelle absolue ou relative) d'inégalités d'accès aux soins obstétricaux. Il suffit ensuite de mener un test d'hétérogénéité entre ces deux mesures pour évaluer si les bénéfices ont été distincts entre les deux groupes et dans quelle direction. Il existe de nombreux tests d'hétérogénéité qui peuvent être menés dans ce contexte. Nous proposons comme exemple le *Cochran Q test* (Kaufman et MacLehose 2013). Ce test peut être conduit sur des mesures absolues ou relatives. Ce test d'hétérogénéité est construit de la manière suivante. Si l'on considère par exemple l'échelle relative et le ratio de prévalence (RP) comme mesure d'association, nous devons estimer un RP qui représente l'effet de l'intervention pour chacun des groupes (ratio de la prévalence d'accès aux soins obstétricaux avant et après l'intervention) RP_1 et RP_2 et pour les deux groupes ensemble (RP_p). Ensuite, il suffit de calculer l'estimé Cochran Q suivant :

$$Cochran's\ Q = \left[\frac{(\beta_1-\beta_P)^2}{VAR(\beta_1)} + \frac{(\beta_2-\beta_P)^2}{VAR(\beta_2)}\right]$$

Dans cette formule, β1 représente le logarithme naturel de RP_1 et β2 représente le logarithme naturel de RP_2 et VAR représente la variance. βp représente le logarithme naturel de RP_p qui est l'effet pour les deux groupes ensemble. Une fois que cet estimé Cochran Q est calculé, il suffit de mener un test de χ^2 (Kaufman et MacLehose 2013) avec un nombre de degrés de liberté égal au nombre de strates moins un.

Ainsi, si RP_1 représente l'effet de l'intervention sur le premier quintile

de l'indice de richesse (le plus pauvre) et RP_2 l'effet de l'intervention sur le cinquième quintile (le plus riche), et RP_p pour les deux groupes ensemble, on estimera que l'intervention est équitable si : i) il y a présence d'un effet hétérogène de l'implantation de l'exemption des frais de soins sur l'accès aux soins obstétricaux et ii) cet effet est plus important pour le premier quintile que pour le cinquième.

Nous avons présenté une approche très simple pour évaluer une potentielle hétérogénéité dans les bénéfices d'une intervention. Cette approche peut être étendue à la prise en compte de facteurs de confusion mesurés dans un contexte d'évaluation des effets d'une intervention. Pour cela, il sera possible de mener, par exemple, des analyses d'appariement (c'est-à-dire le fait de comparer des individus traités et non traités ayant des caractéristiques observables similaires) en utilisant un score de propension (Shadish, Luellen *et al.* 2006). Une fois le score de propension estimé et l'appariement effectué (détails non discutés dans ce chapitre), il sera possible de mener la même approche d'évaluation de l'hétérogénéité via le *Cochran Q test*.

Cependant, dans un contexte non expérimental, il est souvent préférable d'avoir recours à des approches de type quasi-expérimental qui sont présentées en détail dans le chapitre précédent, afin d'évaluer les effets d'une intervention. Dans la section suivante, nous proposons différentes approches permettant d'évaluer l'hétérogénéité des effets d'une intervention en utilisant des méthodes quasi-expérimentales.

Approches quasi-expérimentales

L'utilisation des méthodes quasi-expérimentales s'est largement développée ces dernières années en santé publique puisqu'elles présentent des avantages certains du point de vue de l'inférence causale : elles permettent la prise en compte de facteurs de confusion non mesurés. Cette prise en compte peut parfois avoir une importance considérable en comparaison avec les approches de type avant/après, allant jusqu'à modifier la direction de l'association représentant l'effet d'une exposition (Grabich, Robinson *et al.* 2015). Il existe plusieurs méthodes, par exemple la méthode dite de différence dans les différences (DD) et l'approche de discontinuité dans la régression qui est un cas particulier d'analyse de variable

instrumentale, toutes deux présentées dans le chapitre précédent. Ici nous allons nous concentrer sur l'approche de type DD.

Hormis l'évaluation des effets sur une population dans son ensemble (Benach, Malmusi *et al.* 2011) ou sur un sous-groupe de la population comme les femmes, l'évaluation de l'équité *via* l'utilisation de méthodes quasi-expérimentales demeure rare. Ici, nous proposons trois approches distinctes permettant d'évaluer l'hétérogénéité dans la distribution des effets d'une intervention avec l'exemple de l'approche DD. Pour illustrer, de manière délibérément simplifiée, comment serait organisée l'évaluation de l'exemption du paiement des frais de soins sur l'accès aux soins obstétricaux en utilisant une approche de type DD, nous pouvons considérer la situation suivante. Entre 2000 et 2010, nous mesurons chaque année la prévalence de l'accès aux soins obstétricaux pour des femmes au sein de deux différents pays (pays A et B) similaires. En 2005, le pays A décide d'organiser l'exemption du paiement des frais de soins (exemption de la tarification des accouchements) tandis que le pays B continue d'appliquer des frais pour les accouchements. On suppose que les pays sont échangeables (voir chapitre précédent pour cet aspect) et que le pays B constitue un contrefactuel valide pour représenter ce qui se serait passé dans le pays A sans la mise en place de l'intervention. L'effet causal de l'intervention sur l'accès aux soins obstétricaux sera mesuré par la comparaison de la différence entre les deux pays après 2005 avec la différence entre les deux pays avant 2005. Nous rappelons que cet effet est mesuré dans les modèles de régression par un terme d'interaction entre une variable qui indique la période post-intervention et une variable qui indique le pays avec intervention (Dimick et Ryan 2014). Pour évaluer la répartition des effets entre différents sous-groupes, nous considérons les quintiles de l'indice de richesse.

Analyses stratifiées combinées à un test d'hétérogénéité.

Une première approche consisterait à mener l'analyse de type DD séparément pour les femmes appartenant au premier quintile de richesse et pour les femmes du cinquième quintile. On obtiendrait deux estimations DD de l'effet de l'intervention : une pour les femmes les plus pauvres et une pour les femmes les plus riches. Ensuite, il suffit de mener un test d'hétérogénéité (tel que décrit précédemment) pour évaluer si l'effet de l'intervention sur

l'accès aux soins obstétricaux est hétérogène entre les deux groupes[3]. Si cet effet est hétérogène entre les groupes, d'une part, et que, d'autre part, il est plus important au sein du premier quintile de richesse, alors il sera possible de conclure que la répartition des bénéfices de cette intervention est équitable

Analyses avec double interaction ou analyses de type différences dans les différences dans les différences (DDD)

Une deuxième approche consisterait à évaluer une double interaction. Tel que rappelé précédemment, l'estimation de l'effet d'une intervention via une analyse de type DD se fait au travers d'un terme d'interaction. Ici, il s'agirait d'évaluer la présence d'une double interaction du terme DD avec un indicateur du quintile de richesse. Dit autrement, il s'agit d'évaluer sur le terme DD varie en fonction du niveau de richesse. Ici, l'équité sera évaluée par la significativité de cette interaction et sa direction. Dans notre exemple, comme le premier quintile représente le niveau de richesse le plus bas, ce terme d'interaction doit être négatif pour que la répartition des bénéfices de cette intervention soit jugée comme équitable. Cette approche peut s'apparenter à une analyse de différences dans les différences dans les différences (DDD). Il est possible de trouver des exemples de cette approche dans les écrits en santé publique, avec une approche individuelle (Harper, Strumpf *et al.* 2014) ou une approche mixée à des analyses de séries temporelles (Benmarhnia, Bailey Z *et al.* 2016).

Analyses avec méthodes de décomposition

Dans la section précédente ont été présentées différentes mesures d'inégalités, allant des approches simples comparant deux groupes, aux approches comparant plusieurs groupes (ordonnés ou non) comme l'indice de concentration par exemple.

3. Ce test peut être mené sur plusieurs groupes également, à condition d'adapter le Cochran Q test.

Les méthodes de décomposition ont été développées dans le champ de l'économétrie et initialement utilisées pour expliquer, par exemple, les écarts de revenus selon le sexe (par exemple selon l'âge, le secteur d'activité, le niveau d'expérience). Ces méthodes peuvent être appliquées à la comparaison entre deux groupes où la méthode dite d'Oaxaca Blinder (Oaxaca 1973; O'Donnell et Wagstaff 2008) pourra être utilisée. Ces méthodes peuvent également être appliquées à des mesures d'inégalités plus avancées comme l'indice de concentration (O'Donnell et Wagstaff 2008) tel que discuté plus haut.

Toutefois, leur application à l'évaluation des effets et au travers des méthodes quasi-expérimentales demeure rare également. Ici, nous allons décrire simplement comment appliquer ces méthodes de décomposition à une analyse de type DD.

Les méthodes de décomposition consistent à modéliser quelle serait la diminution des inégalités entre deux ou plusieurs groupes si des expositions étaient ramenées à des niveaux équivalents. Ainsi, dans notre contexte, il suffira de mener des analyses de décomposition entre deux ou plusieurs groupes de richesse par exemple et d'analyser particulièrement le terme d'interaction DD. L'interprétation de ce terme à l'issue de l'analyse de décomposition indiquera si les bénéfices de l'intervention contribuent à réduire les écarts entre les groupes de richesse ainsi que l'amplitude et la direction de cette réduction, ce qui permettra d'évaluer si la répartition des bénéfices de cette intervention peut être jugée comme équitable.

Conclusion

Dans ce chapitre, nous avons présenté la notion d'équité en santé publique, ainsi que plusieurs méthodes pour mesurer les inégalités de santé et différentes approches pour évaluer l'hétérogénéité des effets d'une politique publique. Il est important d'avoir à disposition plusieurs approches méthodologiques correspondant à différentes situations de recherche, où l'on pourra par exemple être intéressé parfois par la comparaison entre deux groupes ou par la compréhension d'un gradient social vis-à-vis d'un indicateur de santé. Chacune des méthodes proposées peut s'inscrire dans différentes approches de santé publique (Benach, Malmusi *et al.* 2011; Benach, Malmusi *et al.* 2013) quant à la manière d'aborder une intervention

de santé publique, que ce soit via des approches populationnelles ou des approches à universalisme proportionné notamment.

L'évaluation de l'équité en santé publique est importante pour s'assurer que, d'une part, des interventions ne contribuent pas à augmenter les inégalités de santé lorsqu'il s'agit d'approches populationnelles, et, d'autre part, qu'elles répondent aux objectifs énoncés dans le cadre d'interventions visant à réduire les inégalités de santé, notamment en permettant que les bénéfices soient plus importants au sein des sous-groupes ou territoires qui sont identifiés comme les plus vulnérables ou dont les besoins sont les plus importants.

Naturellement, cette évaluation n'est qu'une étape parmi d'autres dans la compréhension de la distribution des effets d'une politique publique et des inégalités de santé plus généralement. Avant toute évaluation de l'équité, il est nécessaire de bien connaitre le contexte d'intervention afin de formuler clairement des hypothèses du point de vue de l'hétérogénéité des effets d'une politique publique. Ensuite, il est également important d'étudier les mécanismes qui entrent en jeu dans la variabilité des bénéfices des interventions de santé publique, afin de maintien (?), à la modification de politiques existantes et la formulation d'expérimentations visant la réduction des inégalités de santé.

Références clés

O'Donnell, O. A. et A. Wagstaff (2008). *Analyzing health equity using household survey data: a guide to techniques and their implementation*, World Bank Publications.

Cet ouvrage en libre accès (https://openknowledge.worldbank.org/handle/10986/6896) est un excellent moyen de découvrir les différentes approches de mesures des inégalités de santé, ainsi que les méthodes permettant de comprendre les déterminants de ces inégalités et de leurs changements au cours du temps, notamment en lien avec une intervention spécifique.

Harper, S., N. B. King, S. C. Meersman, M. E. Reichman, N. Breen and J. Lynch (2010). Implicit value judgments in the measurement of health inequalities. *Milbank Quarterly* 88(1): 4-29.

Cet article est incontournable au sein de la littérature des inégalités de santé dans la mesure où, au fil de plusieurs exemples, il met en évidence l'importance du choix de la méthode de mesure des inégalités sur les conclusions. Il détaille également les valeurs implicites qui peuvent guider le choix de certaines méthodes plutôt que d'autres.

Benach, J., D. Malmusi, Y. Yasui, J. M. Martínez and C. Muntaner (2011). « Beyond Rose's strategies: a typology of scenarios of policy impact on population health and health inequalities. » *International Journal of Health Services* 41(1): 1-9.

Cet article propose de résumer les différentes approches permettant de structurer des interventions en santé publique, avec une perspective historique intéressante et une évaluation systématique des impacts de chacune de ces approches sur les inégalités de santé.

Références

Althabe, F., Bergel, E., Cafferata, M. L., Gibbons, L., Ciapponi, A., Alemán, A., ... et Palacios, A. R. (2008). Strategies for improving the quality of health care in maternal and child health in low-and middle-income countries: an overview of systematic reviews. *Paediatric and perinatal epidemiology*, 22(Suppl. 1), 42-60.

Benach, J., Malmusi, D., Yasui, Y. et Martínez, J. M. (2013). A new typology of policies to tackle health inequalities and scenarios of impact based on Rose's population approach. *Journal of epidemiology and community health*, 67(3), 286-291.

Benach, J., Malmusi, D., Yasui, Y., Martínez, J. M. et Muntaner, C. (2011). Beyond Rose's strategies: a typology of scenarios of policy impact on population health and health inequalities. *International Journal of Health Services*, 41(1), 1-9.

Benmarhnia, T., Bailey, Z., Auger, N., Kaiser, D., King, N. et K. JS. (2016). A difference-in-differences approach to assess the effect of a heat action plan on mortality and equity in Montreal, Quebec. *Environmental Health Perspectives*, 124(11), 1694-1699.

Benmarhnia, T., Grenier, P., Brand, A., Fournier, M., Deguen, S. et Smargiassi, A. (2015). Quantifying vulnerability to extreme heat in time series analyses:

a novel approach applied to neighborhood social disparities under climate change. *International Journal of Environmental Research and Public Health*, 12(9), 11869-11879.

Benmarhnia, T., Rey, L., Cartier, Y., Clary, C. M., Deguen, S. et Brousselle, A. (2014). Addressing equity in interventions to reduce air pollution in urban areas: a systematic review. *International journal of public health*, 59(6), 933-944.

Blas, E. et Kurup, A. S. (2010). *Equity, social determinants and public health programmes*. Geneva: World Health Organization.

Braveman, P. et Gruskin, S. (2003). Defining equity in health. *Journal of epidemiology and community health*, 57(4), 254-258.

Braveman, P. A., Cubbin, C., Egerter, S., Chideya, S., Marchi, K. S., Metzler, M. et Posner, S. (2005). Socioeconomic status in health research: one size does not fit all. *Jama*, 294(22), 2879-2888.

Brown, T., Platt, S. et Amos, A. (2014). Equity impact of interventions and policies to reduce smoking in youth: systematic review. *Tobacco control*, 23(e2), e98-e105.

Brown, T., Platt, S. et Amos, A. (2014). Equity impact of population-level interventions and policies to reduce smoking in adults: a systematic review. *Drug and alcohol dependence*, 138, 7-16.

Cartier, Y., Benmarhnia, T. et Brousselle, A. (2015). Tool for assessing health and equity impacts of interventions modifying air quality in urban environments. *Evaluation and program planning*, 53, 1-9.

Daniels, N. (2006). Equity and population health: toward a broader bioethics agenda. *Hastings Center Report*, 36(4), 22-35.

Dimick, J. B. et Ryan, A. M. (2014). Methods for evaluating changes in health care policy: the difference-in-differences approach. *Jama*, 312(22), 2401-2402.

Dzakpasu, S., Powell-Jackson, T. et Campbell, O. M. (2013). Impact of user fees on maternal health service utilization and related health outcomes: a systematic review. *Health policy and planning*, 29(2), 137-150.

Ellis, R. P. (1987). The revenue generating potential of user fees in Kenyan government health facilities. *Social science et medicine*, 25(9), 995-1002.

Farrelly, M. C., Nonnemaker, J. M., Chou, R., Hyland, A., Peterson, K. et Bauer,

U. E. (2005). Changes in hospitality workers' exposure to secondhand smoke following the implementation of New York's smoke-free law. *Tobacco Control*, 14(4), 236-241.

Filmer, D. et Pritchett, L. (1999). The effect of household wealth on educational attainment: evidence from 35 countries. *Population and development review*, 25(1), 85-120.

Frazer, K., Callinan, J. E., McHugh, J., van Baarsel, S., Clarke, A., Doherty, K. et Kelleher, C. (2016). Legislative smoking bans for reducing harms from secondhand smoke exposure, smoking prevalence and tobacco consumption. *Cochrane Database of Systematic Reviews*, (2), p.-p.

Gilman, S. E., Rende, R., Boergers, J., Abrams, D. B., Buka, S. L., Clark, M. A., ... et Lipsitt, L. P. (2009). Parental smoking and adolescent smoking initiation: an intergenerational perspective on tobacco control. *Pediatrics*, 123(2), e274-e281.

Grabich, S. C., Robinson, W. R., Engel, S. M., Konrad, C. E., Richardson, D. B. et Horney, J. A. (2015). County-level hurricane exposure and birth rates: application of difference-in-differences analysis for confounding control. *Emerging themes in epidemiology*, 12(1), 1.

Hajizadeh, M., Campbell, M. K. et Sarma, S. (2014). Socioeconomic inequalities in adult obesity risk in Canada: trends and decomposition analyses. *The European Journal of Health Economics*, 15(2), 203-221.

Harper, S., King, N. B., Meersman, S. C., Reichman, M. E., Breen, N. et Lynch, J. (2010). Implicit value judgments in the measurement of health inequalities. *The Milbank Quarterly*, 88(1), 4-29.

Harper, S., Lynch, J., Meersman, S. C., Breen, N., Davis, W. W. et Reichman, M. E. (2008). An overview of methods for monitoring social disparities in cancer with an example using trends in lung cancer incidence by area-socioeconomic position and race-ethnicity, 1992–2004. *American journal of epidemiology*, 167(8), 889-899.

Harper, S., Strumpf, E. C., Burris, S., Smith, G. D. et Lynch, J. (2014). The effect of mandatory seat belt laws on seat belt use by socioeconomic position. *Journal of policy analysis and management*, 33(1), 141-161.

Hoffman, S. J. et Tan, C. (2015). Overview of systematic reviews on the health-related effects of government tobacco control policies. *BMC public health*, 15(1), 744-754.

Hosseinpoor, A. R., Van Doorslaer, E., Speybroeck, N., Naghavi, M., Mohammad, K., Majdzadeh, R., ... et Vega, J. (2006). Decomposing socioeconomic inequality in infant mortality in Iran. *International journal of epidemiology*, 35(5), 1211-1219.

Houweling, T. A., Kunst, A. E., Huisman, M. et Mackenbach, J. P. (2007). Using relative and absolute measures for monitoring health inequalities: experiences from cross-national analyses on maternal and child health. *International journal for equity in health*, 6(1), 1.

Howe, L. D., Hargreaves, J. R., Gabrysch, S. et Huttly, S. R. (2009). Is the wealth index a proxy for consumption expenditure? A systematic review. *Journal of epidemiology and community health*, 63(11), 871-877.

Kakwani, N., Wagstaff, A. et Van Doorslaer, E. (1997). Socioeconomic inequalities in health: measurement, computation, and statistical inference. *Journal of econometrics*, 77(1), 87-103.

Kaufman, J. S. et MacLehose, R. F. (2013). Which of these things is not like the others?. *Cancer*, 119(24), 4216-4222.

Kawachi, I. et Kennedy, B. P. (1997). The relationship of income inequality to mortality: does the choice of indicator matter?. *Social science et medicine*, 45(7), 1121-1127.

Kawachi, I., Subramanian, S. et Almeida-Filho, N. (2002). A glossary for health inequalities. *Journal of epidemiology and community health*, 56(9), 647-652.

Lagarde, M. et Palmer, N. (2008). The impact of user fees on health service utilization in low-and middle-income countries: how strong is the evidence?. *Bulletin of the World Health Organization*, 86(11), 839-848C.

Litvack, J. I. et Bodart. C. (1993). User fees plus quality equals improved access to health care: results of a field experiment in Cameroon. *Social science et medicine*, 37(3), 369-383.

Lynch, J., Smith, G. D., Harper, S. et Bainbridge, K. (2006). Explaining the social gradient in coronary heart disease: comparing relative and absolute risk approaches. *Journal of epidemiology and community health*, 60(5), 436-441.

Mackenbach, J. P. et Kunst, A. E. (1997). Measuring the magnitude of socio-economic inequalities in health: an overview of available measures

illustrated with two examples from Europe. *Social science et medicine*, 44(6), 757-771.

Marmot, M., Friel, S., Bell, R., Houweling, T. A., Taylor, S. et Commission on Social Determinants of Health. (2008). Closing the gap in a generation: health equity through action on the social determinants of health. *The Lancet*, 372(9650), 1661-1669.

McKinnon, B., Harper, S. et Kaufman, J. S. (2015). Who benefits from removing user fees for facility-based delivery services? Evidence on socioeconomic differences from Ghana, Senegal and Sierra Leone. *Social Science et Medicine*, 135, 117-123.

McKinnon, B., Harper, S., Kaufman, J. S. et Bergevin, Y. (2014). Removing user fees for facility-based delivery services: a difference-in-differences evaluation from ten sub-Saharan African countries. *Health policy and planning*, 30(4), 432-441.

O'Donnell, O. A. et Wagstaff, A. (2008). *Analyzing health equity using household survey data: a guide to techniques and their implementation*. Washington D.C.: World Bank Publications.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International economic review*, 693-709.

Pamuk, E. R. (1985). Social class inequality in mortality from 1921 to 1972 in England and Wales. *Population studies*, 39(1), 17-31.

Pearcy, J. N. et Keppel, K. G. (2002). A summary measure of health disparity. *Public health reports*, 117(3), 273-280.

Regidor, E. (2004). Measures of health inequalities: part 2. *Journal of epidemiology and community health*, 58(11), 900-903.

Ridde, V. et Morestin, F. (2011). A scoping review of the literature on the abolition of user fees in health care services in Africa. *Health policy and planning*, 26(1), 1-11.

Russell, S. et Gilson, L. (1997). User fee policies to promote health service access for the poor: a wolf in sheep's clothing?. *International Journal of Health Services*, 27(2), 359-379.

Say, L. et Raine, R. (2007). A systematic review of inequalities in the use of maternal health care in developing countries: examining the scale of

the problem and the importance of context. *Bulletin of the World Health Organization*, 85(10), 812-819.

Shadish, W. R., Luellen, J. K. et Clark, M. (2006). Propensity scores and quasi-experiments: A testimony to the practical side of lee sechrest.

Smith, M. J. (2015). Health equity in public health: Clarifying our commitment. *Public Health Ethics*, 8(2), 173-184.

Starfield, B. (2007). Pathways of influence on equity in health. *Social Science et Medicine*, 64(7), 1355-1362.

Theil, H. (1967). *Economics and information theory*.

Wagstaff, A., Van Doorslaer, E. et Paci, P. (1991). On the measurement of horizontal inequity in the delivery of health care. *Journal of health economics*, 10(2), 169-205.

Wagstaff, A., Van Doorslaer, E. et Watanabe, N. (2003). On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam. *Journal of econometrics*, 112(1), 207-223.

Whitehead, M. (1992). The concepts and principles of equity and health. *International journal of health services*, 22(3), 429-445.

Yates, R. (2009). Universal health care and the removal of user fees. *The Lancet*, 373(9680), 2078-2081.

Yoder, R. A. (1989). Are people willing and able to pay for health services? *Social science et medicine*, 29(1), 35-42.

Résumé / Abstract / Resumen

L'équité en santé est largement identifiée comme un objectif auquel les politiques et activités de santé publique doivent répondre en priorité, mais dont la définition varie considérablement. Dans ce chapitre, nous proposons une définition de l'équité en santé qui fait référence à la distribution des bénéfices des politiques publiques. Dans ce chapitre nous présentons différentes méthodes de mesures des inégalités de santé, leur application à l'évaluation des effets des interventions à travers l'exemple de l'effet de l'exemption du paiement des frais de soins sur l'accès aux soins obstétricaux dans le cadre d'un exemple simulé. Ces mesures incluent notamment l'indice de concentration et l'indice de Theil. Enfin, nous proposons différentes

approches pour les appliquer à l'évaluation des effets des interventions donc à l'équité incluant une approche simple et une base sur l'utilisation de méthodes quasi-expérimentales. L'évaluation de l'équité en santé publique est importante pour s'assurer que, d'une part, des interventions ne contribuent pas à augmenter les inégalités de santé et qu'elles répondent aux objectifs énoncés dans le cadre d'interventions visant à réduire les inégalités de santé.

***

Health equity is considered a priority for public health interventions and policies, but the concept of an equitable intervention varies in definition. We define an equitable intervention as one that has heterogeneity in its effects, and provides greater benefits for subgroups or territories of a population that have been defined as vulnerable. In this chapter, we assess the impact of an intervention establishing free obstetric care (childbirth costs excluded) on increasing equity and access to care. We notably present the concentration index and the Theil Index. Simulated data from population surveys is used to discuss important issues regarding measurement of inequalities and different indicators that can be used in evaluating the equitability of an intervention. Lastly, two evaluation strategies are discussed: a simple evaluation approach to illustrate the application of inequality measurement methods and an approach based on quasi-experimental designs. These tools to assess public health equity is important to ensure that interventions meet objectives to reduce inequalities and prevent them from contributing to increasing health disparities.

***

La equidad en salud es ampliamente identificada como un objetivo al que las políticas y actividades de salud pública deben responder como una prioridad, pero cuya definición varía considerablemente. En este capítulo proponemos una definición de equidad en salud que se refiere a la distribución de los beneficios de las políticas públicas. En este capítulo se presentan diferentes métodos para medir las desigualdades en salud, su aplicación a la evaluación de los efectos de las intervenciones a través del ejemplo del efecto de la exención de las tasas de atención sanitaria sobre el acceso a la atención obstétrica en un ejemplo simulado. Estas medidas incluyen el índice de concentración y el índice de Theil. Finalmente, proponemos diferentes enfoques para aplicarlos a la evaluación de los

efectos de las intervenciones y, por lo tanto, a la equidad, incluyendo un enfoque simple y una base sobre el uso de métodos cuasi-experimentales. La evaluación de la equidad en salud pública es importante para garantizar que, por una parte, las intervenciones no contribuyan a aumentar las desigualdades en salud y, por otra parte, que cumplan los objetivos establecidos en las intervenciones para reducir las desigualdades en salud.

***

Tarik Benmarhnia est Professeur assistant d'épidémiologie à l'Université de Californie à San Diego au sein de l'Institut d'océanographie Scrippset de l'école de médecine. Il a complété son doctorat en épidémiologie à l'Université de Montréal et à l'Université Paris Sud. Il est également chercheur associé au Chili et en France et membre du comité éditorial de *Environmental Health Perspectives*. Sa recherche porte sur les impacts des évènements extrêmes sur la santé avec un focus sur les populations vulnérables et les implications en termes de politiques publiques. Il développe également des approches méthodologiques pour évaluer l'effet des politiques environnementales sur la santé comme les mesures d'adaptation aux changements climatiques et les mesures de contrôle de la pollution atmosphérique.

Britt McKinnon est chercheuse au centre de Santé mondiale à l'hôpital des Enfants malades et professeure assistante à l'École de santé publique Dalla Lana de l'Université de Toronto. Elle a obtenu son doctorat en épidémiologie à l'université McGill et fait un postdoctorat au sein de L'Institut des politiques sociales et de Santé de McGill. Sa recherche se concentre en en santé mondiale sur la santé maternelle et des enfants, en épidémiologie sociale et l'évaluation des politiques publiques et programmes en santé.

Citation

Tarik Benmarhnia et Britt McKinnon. (2019). Les mesures de l'équité. Un exemple d'intervention de gratuité des soins obstétricaux. In *Évaluation des interventions de santé mondiale. Méthodes avancées*. Sous la direction de Valéry Ridde et Christian Dagenais, pp. 283-309. Québec : Éditions science et bien commun et Marseille : IRD Éditions.

12. L'analyse coût-efficacité

Une intervention de décentralisation des soins VIH/SIDA à Shiselweni, Swaziland

GUILLAUME JOUQUET

L'analyse coût-efficacité

Définition de la méthode

Étude rétrospective de coût-efficacité du point de vue du prestataire de service avec calcul du ratio de coût efficacité incrémental (ICER)

Forces de la méthode

- compare deux alternatives en intégrant à la fois des aspects qualitatifs (efficacité) et économiques (coût) des programmes de santé
- repose sur des données réelles d'un programme grandeur nature

Défis de la méthode

- l'accès nécessaire aux données financières des prestataires de service peut être un défi
- la mesure du gain d'efficacité liée à l'approche alternative en isolant les éventuels facteurs de confusion.

Chaque fois qu'on me demande de faire une évaluation économique, ma première question est de savoir à qui et à quoi cela va servir. Est-ce pour justifier des dépenses vis-à-vis de tiers, pour aider à prendre une décision relative à l'allocation de ressources, pour simuler de futures dépenses ou pour faire du plaidoyer ? La réponse à cette question permettra de déterminer quel type d'évaluation économique sera le plus approprié. Il en existe trois principaux types :

- L'évaluation du coût total, qui consiste à calculer le coût de l'ensemble des intrants du programme (par exemple : les ressources humaines et l'approvisionnement médical)
- L'évaluation de la productivité, qui se base sur les « quantités produites ». Il s'agira de diviser le coût total par les extrants du programme (par

exemple : le nombre de patient-e-s traité-e-s).
- L'étude de coût efficacité, qui introduit une notion qualitative en divisant le coût total par une mesure du résultat du programme (par exemple : le nombre de patient-e-s guéri-e-s)

L'évaluation pourra être conduite selon différents points de vue :

- le point de vue du prestataire de service (par exemple : l'ONG et/ou le ministère de la Santé qui supportent le coût du programme)
- le point de vue du bénéficiaire (par exemple : le ou la patient-e qui doit payer des frais de transport pour se rendre dans un centre de soins).

La démarche la plus complète consiste, dans la mesure du possible, à prendre en considération ces deux points de vue.

Les deux principales méthodes que nous pouvons utiliser, de façon exclusive ou complémentaire, pour évaluer les coûts sont la méthode des coûts complets et la méthode des ingrédients.

- La méthode des coûts complets est une méthode descendante (Conteh 2004) : on part de l'ensemble des dépenses enregistrées en comptabilité que l'on va catégoriser en coûts directs et indirects. Les coûts directs seront imputés directement à des centres de coûts ou à des activités, alors que les coûts indirects seront alloués en fonction d'une clé de répartition.
- La méthode des ingrédients est une méthode ascendante (WHO-CHOICE 2016) : il s'agit pour chaque intrant utilisé dans une activité donnée d'en déterminer la quantité nécessaire et le coût unitaire. Le coût de l'activité est alors obtenu en faisant la somme des produits du coût unitaire par la quantité nécessaire de chaque intrant.

Le choix du type d'évaluation, des points de vue et des méthodes dépendra non seulement des objectifs de l'étude, mais aussi d'une double contrainte (Ridde et Dagenais 2012, p172) :

- le temps et les ressources allouées à l'évaluation
- la difficulté d'accès à des données fiables et exploitables.

Dans le cas présent, le client de l'évaluation est Médecins sans Frontière

(MSF), une organisation non gouvernementale (ONG) médicale internationale. Cette ONG a mis en place en partenariat avec le Ministère de la Santé publique (MSP) du Swaziland un programme de traitement du syndrome d'immunodéficience acquise (SIDA). Ce programme a la particularité d'être décentralisé et intégré au niveau des centres de santé primaire (CSP). Il s'agit d'une démarche innovante au Swaziland où le traitement du SIDA est centralisé au niveau des centres de santé secondaire (CSS). Les approches décentralisées du traitement du SIDA dans les pays en développement ont fait l'objet de plusieurs études montrant leur efficacité (Bedelu *et al.* 2007; Cohen *et al.* 2009; Loubiere *et al.* 2009; Bemelmans *et al.* 2010; Chan *et al.* 2010,). Le but de la présente évaluation est d'y ajouter un volet économique. Pour comparer une intervention médicale avec une autre ayant le même objectif, le type d'évaluation économique le plus approprié est une étude de coût-efficacité avec calcul du ratio de coût-efficacité incrémental (ICER) (Drummond 2005, p39). Ce ratio indiquera le coût supplémentaire à payer pour gagner une année de vie supplémentaire dans le programme le plus efficace par rapport au programme le moins efficace. Le mode de calcul et l'interprétation de ce ratio sont expliqués dans la quatrième partie de ce chapitre.

Une approche médicale décentralisée facilite l'accès aux soins pour les patient-e-s, mais rend la tâche plus difficile pour le prestataire de service (Mulamba *et al.* 2010). Pour appréhender cette évaluation dans sa globalité, il conviendrait donc de la réaliser du double point de vue du patient-e et du prestataire de service. Cependant, en raison des contraintes de ressources et de temps alloués par notre client, nous avons préféré nous limiter au point de vue des prestataires de service, en l'occurrence MSF et le MSP.

Le calcul de l'ICER implique d'estimer le coût et l'efficacité de chaque approche. Pour estimer les coûts économiques et financiers, nous avons opté pour la méthode des ingrédients qui présente certains avantages, mais aussi certaines limites que nous exposerons plus en détail dans la cinquième partie de ce chapitre. Pour la mesure de l'efficacité, nous avons choisi le maintien dans la thérapie antirétrovirale (ARV) à 12 mois comme étant l'effet attendu de l'approche. La raison de ce choix est détaillée dans la quatrième partie de ce chapitre (mesure de l'efficacité).

Les principales difficultés pour une évaluation telle que celle-ci résident dans :

- l'accès aux données financières des prestataires de services
- la mesure du gain d'efficacité liée à l'approche alternative en isolant les éventuels facteurs de confusion.

Contexte et programme évalué

Swaziland

Le Swaziland, avec une population d'environ 1 million de personnes sur 17 000 km², est un petit pays montagneux enclavé entre l'Afrique du Sud et le Mozambique. C'est l'une des trois monarchies restantes du continent africain et la seule où le roi a encore un pouvoir exécutif. Dans les années 1980 et au début des années 1990, le royaume a connu un développement économique qui a profité de façon inégale à la population. Le Swaziland est en effet le septième pays le plus inégalitaire du monde et on estime que 81 % de la population vit avec moins de 2 dollars par jour. Ce développement économique a quand même permis à la situation sanitaire de la population Swazi de considérablement s'améliorer jusqu'au début des années 1990 (UNDP 2009). Malheureusement, l'arrivée du SIDA a complètement changé la situation. Le pays présente la plus forte prévalence au monde du VIH (virus de l'immunodéficience humaine) parmi les adultes âgés de 15 à 49 ans : 31 % d'entre eux et elles sont infecté-e-s et risquent de mourir dans les années à venir à défaut de recevoir un traitement antirétroviral (ARV). Le Swaziland a également l'un des taux les plus élevés d'incidence de la tuberculose (WHO 2009). La co-épidémie a sensiblement contribué à une réduction de moitié de l'espérance de vie dans les deux dernières décennies ; celle-ci est passée de 60 ans dans les années 1990 à 31 en 2007 (UNAIDS 2010).

Région de Shiselweni

En 2007, après une évaluation de quatre mois dans tout le pays, MSF décida de lancer, en partenariat avec le MSP, un programme de prévention et de traitement du SIDA et de la tuberculose dans la région de Shiselweni. Un protocole d'accord fut signé en novembre de la même année. La région de Shiselweni fut choisie par MSF, car elle est considérée comme étant la plus pauvre et la plus reculée. Elle se distingue entre autres par des

infrastructures de santé insuffisantes et de mauvais résultats de santé publique. Il s'agit d'une région montagneuse et rurale divisée en trois zones de santé : Nhlangano, Hlatikulu et Malangeni.

Chaque zone de santé comporte une structure centrale de soins de santé secondaire (CSS) et entre sept et huit centres de soins primaires (CSP) en périphérie (22 au total sur la région). Les principaux problèmes structurels identifiés concernant les soins VIH/TB sont :

- la pénurie chronique de personnel
- la centralisation et l'absence d'intégration des services VIH/TB
- une capacité insuffisante à fournir des soins du VIH/TB dans les CSP.

Ajoutons à cela le paysage montagneux, la dispersion des populations et les coûts élevés de transport. Tout ceci fait que les soins VIH/TB étaient inaccessibles pour la majorité des Swazis vivant dans les zones rurales de la région.

Le programme MSF

Dans ce contexte de ressources limitées et d'accès difficile aux soins, l'objectif du programme mis en place par MSF est de « soutenir le ministère de la Santé dans l'extension des services pour combler le manque de soins de santé pour les personnes vivant avec le VIH (PVVIH) dans la région de Shiselweni ».

Les résultats attendus à court, moyen et long termes sont :

- dépister les PVVIH ayant besoin d'un traitement
- réduire les barrières géographiques de l'accès aux soins
- améliorer le maintien des patient-e-s dans une thérapie antirétrovirale
- augmenter la durée de vie des patient-e-s et diminuer le risque de transmission du VIH.

Pour atteindre ces résultats, la stratégie de MSF a consisté à favoriser la décentralisation et l'intégration des soins de santé VIH au niveau des CSP dont il a fallu revoir l'organisation. Une première difficulté était l'insuffisance des ressources humaines de santé pour faire face à la charge de travail supplémentaire liée à la prise en charge des PVVIH. MSF y a donc alloué du

personnel de santé en plus du personnel MSF déjà présent. Une deuxième difficulté résidait dans l'absence de médecins au niveau des CSP. MSF y a donc alloué un médecin mobile et a organisé le transfert vers les infirmières de certaines tâches normalement dévolues aux médecins. Pour que les infirmières puissent absorber cette nouvelle charge de travail, MSF a organisé le recrutement et la formation de patient-e-s expert-e-s qui se sont vu confier certaines tâches normalement dévolues aux infirmières, notamment le soutien psychologique et le suivi de l'adhérence au traitement des patient-e-s. Une troisième difficulté était liée aux suivis de laboratoire nécessaires pour les patient-e-s sous ARV. Il a donc été mis en place dans chaque CSP des mini-laboratoires dotés d'appareils simples d'usage permettant de faire les tests nécessaires. MSF a également recruté et formé des personnels communautaires non qualifiés (les « phlébotomistes ») chargés d'effectuer ces tests au sein des mini laboratoires. Une équipe mobile de supervision a été constituée, comprenant en plus du médecin signalé plus haut, un responsable infirmier, un responsable de laboratoire et un responsable des client-e-s expert-e-s. Enfin, la logistique a été organisée pour que tout cela fonctionne, à savoir le transport de l'équipe mobile de supervision, mais aussi celui des intrants médicaux et de laboratoire ainsi que celui d'échantillons de sang pour certains tests toujours effectués au niveau du laboratoire central.

Plan/devis/protocole d'évaluation

Objectif

En 2012, soit cinq années après le début du programme de Shiselweni, MSF a souhaité analyser cette expérience afin d'en tirer les leçons et de les partager au sein du mouvement MSF ainsi qu'avec les autorités nationales et la communauté internationale. L'évaluation économique décrite ici fait partie d'une évaluation plus globale menée par quatre évaluateurs et évaluatrices indépendantes et commanditée par MSF dans un double but de documentation des expériences acquises et de plaidoyer (Turashvili *et al.* 2014).

Questions d'évaluation

La question de l'évaluation économique était de savoir si l'approche décentralisée du traitement du SIDA mise en place par MSF dans la région de Shiselweni était coût-efficace par rapport à l'approche centralisée en vigueur partout ailleurs au Swaziland. Cette question se justifiait par le fait que la décentralisation des services au niveau des CSP était censée entraîner à la fois un surcoût pour le prestataire de services, mais aussi un moindre coût et de meilleurs résultats de santé du point de vue des bénéficiaires.

Pour répondre à cette question, il aurait été idéal de mener :

- une évaluation de coût-efficacité
- du point de vue du prestataire de services et des bénéficiaires
- comparant une population test et une population témoin en tout point semblables, hormis l'approche de soins VIH adoptée (centralisée ou décentralisée)
- en calculant le ratio de coût-efficacité incrémental (ICER) d'une approche par rapport à l'autre

Considérant les contraintes de ressources, de temps et de disponibilité de l'information, nous avons dû revoir à la baisse ce schéma idéal qui se limita finalement à :

- une évaluation de coût-efficacité basée sur une étude rétrospective de données recueillies systématiquement par le MSP et MSF dans la zone de santé de Nhlangano (limitation géographique)
- du point de vue du prestataire de service (seulement)
- comparant les patient-e-s sous ARV suivis dans les CSP avec les patient-e-s sous ARV suivi dans le CSS de la zone de santé en 2012
- en calculant le ratio de coût-efficacité incrémental (ICER) des patient-e-s CSP par rapport aux patient-e-s CSS

L'étude a été autorisée par le ministère de la Santé. Étant donné que l'analyse était fondée sur les données cliniques de routine, l'approbation d'un comité éthique et le consentement individuel des patient-e-s n'ont pas été demandés. Toutes les informations liées aux patient-e-s ont été traitées de façon anonyme en utilisant des numéros d'identification codés.

Les informations susceptibles de révéler l'identité des patient-e-s n'ont pas été utilisées.

Budget et planification

Cette évaluation économique s'est déroulée entre octobre 2012 et juillet 2013, en trois grandes phases : définition des termes de référence, collecte et analyse des données sur place et rédaction du rapport d'évaluation.

Les principaux coûts de l'étude étaient liés, d'une part, aux ressources humaines (un économiste indépendant pour 40 jours-personne et l'épidémiologiste du programme MSF à Shiselweni pour 20 jours-personne) et, d'autre part, aux frais de voyage pour les deux visites sur place de l'économiste (l'épidémiologiste était quant à elle déjà sur place).

Application de la méthode

Une fois que la question et le devis ont bien été définis en partenariat avec le commanditaire de l'évaluation, on peut distinguer quatre grandes étapes dans le déroulement de la méthode que nous avons suivi :

- Collecte des données
- Estimation du coût
- Mesure de l'efficacité
- Analyse et interprétation des résultats

Collecte des données

La collecte des données doit permettre de calculer le ratio incrémental de coût-efficacité (ICER) entre les deux approches.

L'ICER se calcule de la façon suivante : $ICER = C_{CSP} - C_{CSS}/E_{CSP} - E_{CSS}$. « C » est le coût par année d'un-e patient-e sous ARV suivi dans un centre de santé primaire (CSP) ou secondaire (CSS), « E » est l'efficacité mesurée par la proportion de patient-e-s maintenu-e-s dans la thérapie antirétrovirale à 12 mois dans les CSP ou les CSS.

Il nous a donc fallu collecter des informations qui nous ont permis, pour chaque approche, d'estimer les coûts et de mesurer l'efficacité.

Pour ce faire, nous avons dû croiser plusieurs sources de données quantitatives (bases de données comptable, des ressources humaines, du suivi médical, des approvisionnements et du suivi des véhicules) et qualitatives (rapports d'étude, entretiens) en provenance de MSF ou du MSP. C'est la confrontation de ces différentes sources quantitatives et qualitatives qui nous a permis d'avoir une vue d'ensemble satisfaisante. C'est pourquoi la visite *in situ* était incontournable : elle m'a permis de me faire une idée du type d'information disponible (ou non) et sous quelle forme et de compléter par des entretiens les données manquantes, incohérentes ou difficilement compréhensibles, issues de l'approche quantitative.

Estimation du coût

Le coût a été estimé en utilisant la méthode des ingrédients qui comporte deux éléments clés, soit : 1) identifier les quantités des ingrédients utilisées et 2) établir le coût unitaire ou le prix de chaque ingrédient. Le coût par patient-e a donc été estimé pour chacune des deux approches comme suit :

C = Q1P1 +... + QnPn, avec :

- C = coût par patient-e
- Q1... Qn = quantités de ressources
- P1... PN = prix (ou les coûts unitaires) de chacune des ressources quantifiées.

Autrement dit, le coût par patient-e est obtenu en additionnant le coût de chacun des ingrédients utilisés pour prodiguer les soins nécessaires. Le coût de chaque ingrédient est calculé en multipliant son prix unitaire par la quantité utilisée.

Le coût par patient-e doit être exprimé selon une unité commune de temps afin de comparer un-e patient-e à un-e autre dans les CSP et dans le CSS. Nous avons choisi ici d'estimer les coûts en termes de patient-e/année (un-e patient-e qui, par exemple, serait mort-e trois mois après le début de son traitement ne représente ni le même coût ni la même efficacité qu'un-

e patient-e toujours en vie et sous traitement un an après le début de son traitement).

Pour calculer l'ICER, nul besoin d'estimer les coûts qui sont identiques d'une approche à l'autre. Typiquement dans notre cas, il n'y avait aucune raison pour que le coût des ARV proprement dits diffère entre un-e patient-e suivi-e dans un CSP et un-e patient-e suivi-e dans un CSS. Cependant, nous avons tout de même pris ce coût en considération dans notre analyse pour montrer qu'il représente la majeure partie du coût total pour un-e patient-e sous ARV. Les autres coûts que nous avons considérés sont :

- Le coût des consultations par les personnels de santé
- Le coût du suivi psychologique et d'adhérence par les patient-e-s expert-e-s
- Le coût de l'équipe de supervision
- Le coût des quatre principaux tests de laboratoire requis pour les patient-e-s sous ARV
- Le coût de transport de l'équipe de supervision, des intrants médicaux et de la collecte des échantillons de sang pour certains tests toujours réalisés au niveau central.

Nous avons également choisi d'omettre certains coûts qui, bien que pouvant varier d'une approche à l'autre, restent cependant peu significatifs. Nous pouvons citer notamment l'amortissement des bâtiments et le coût des médicaments pour les infections opportunistes.

Le coût des ARV par patient-e/année a été obtenu sur la base du prix annuel des traitements et de la proportion de patient-e-s sous les différents régimes possibles d'ARV.

Pour calculer le coût des consultations, nous avions besoin de connaitre le nombre de consultations par patient-e/année ainsi que le coût unitaire d'une consultation. Le nombre de consultations par patient-e/année a été obtenu grâce au fichier informatique de suivi individuel des patient-e-s sous ARV au Swaziland (APMR). Pour le coût unitaire d'une consultation, nous savions qu'il serait différent dans un CSS comportant un service ARV dédié et dans un CSP où les soins VIH sont intégrés aux autres soins.

S'agissant du centre ARV du CSS, le calcul était direct puisque ce service ne reçoit pratiquement que des patient-e-s sous ARV (un faible pourcentage concerne des patient-e-s en prétraitement ARV). Il s'agissait alors de

calculer le coût annuel du personnel travaillant dans le service, d'y ajouter un pourcentage pour couvrir les frais de fonctionnement et de diviser le tout par le nombre total de consultations effectuées par le service au cours de l'année.

S'agissant des CSP, les consultations effectuées par une même infirmière peuvent être enregistrées dans trois registres différents : les consultations générales (81 % du total), les consultations ARV (14 % du total) et les consultations anténatales et postnatales (5 % du total). L'analyse mensuelle par CSP des données informatisées de ces trois registres faisait apparaitre quelques incohérences dont certaines pouvaient être corrigées facilement. Il manquait par exemple parfois un mois de données, ce qui était facilement visible et expliqué après enquête par une absence de la personne qui effectue les saisies. Nous avons donc corrigé en approximant le mois manquant par la moyenne du mois d'avant et du mois d'après. Deux des sept CSP analysés présentaient de telles incohérences que nous avons préféré écarter de l'analyse. Nos investigations nous ont confirmé que les encodeurs des deux CSP en question ne fournissaient pas un travail satisfaisant. Nous avons donc calculé le coût moyen d'une consultation sur la base des trois registres annuels des cinq CSP retenus, soit 91 183 consultations au total. Le coût moyen d'une consultation a alors été estimé selon la même méthode que pour le service ARV du CSS. Nous avons alors considéré que le temps passé par le personnel de santé pour chaque type de consultation était normalement distribué et que le coût moyen de l'ensemble des consultations était un bon indicateur du coût moyen des consultations ARV.

Pour estimer le coût par patient-e/année des patient-e-s expert-e-s en charge du soutien psychologique et du suivi de l'adhérence, la tâche était plus aisée, car ces personnels sont dédiés à 100 % aux patient-e-s sous ARV. Il suffisait donc de diviser leur coût annuel par le nombre de patient-e-s/années dans chaque type de centre de santé.

Nous avons considéré comme négligeable le coût de la supervision du service ARV du CSS, car il est presque inexistant. En revanche, nous avons calculé le coût du personnel de l'équipe mobile dédiée aux soins VIH dans les CSP et constituée d'un médecin traitant, d'un superviseur infirmier, d'un superviseur des patient-e-s expert-e-s et d'un responsable de laboratoire. Comme la zone de supervision de cette équipe couvre plus de CSP que ceux de notre échantillon, nous avons effectué une règle de trois pour estimer le

coût s'y rapportant spécifiquement, coût que nous avons ensuite divisé par le nombre de patient-e-s/années de notre échantillon.

Le coût unitaire de chaque test diffère forcément entre un-e patient-e suivi dans un CSP et un-e patient-e suivi-e dans un CSS, car les technologies et les catégories de personnel ne sont pas les mêmes dans un mini-laboratoire ou au sein du laboratoire central. Pour estimer le coût de ces tests, nous nous sommes basés sur l'entrevue de la personne responsable des laboratoires à MSF. Nous avons constitué ensemble la liste et les quantités de réactifs et consommables nécessaires pour chaque test dans chaque type de laboratoire (CSP/CSS). Nous avons complété cette entrevue par des données en provenance du service des achats de MSF pour établir le prix unitaire de chaque réactif et consommable. Le coût des personnels de laboratoire a été obtenu en divisant le coût annuel de ces personnels par le nombre de tests annuels effectués au sein de chaque entité. Pour estimer le nombre de tests effectués par patient-e et par année, nous nous sommes basés sur le protocole en vigueur, car les données de l'APMR se sont avérées peu fiables sur la question.

Le dernier type de coût que nous avons pris en compte est celui du transport fourni par MSF pour les équipes de supervision, les intrants médicaux et certains échantillons de sang. Nous avons calculé, grâce au suivi comptable et au suivi des véhicules de MSF, un coût par kilomètre comprenant le chauffeur, l'amortissement du véhicule, le fuel, l'entretien et les réparations. Le fichier de suivi des véhicules MSF nous a également permis d'estimer le nombre de kilomètres parcourus pour desservir les CSP de notre échantillon.

Toutes les données relatives aux coûts ont été enregistrées sous Excel et analysées grâce aux tableaux croisés dynamiques.

Mesure de l'efficacité

Pour mesurer l'efficacité des deux approches, nous avons choisi comme indicateur le maintien des patient-e-s dans la thérapie antirétrovirale à 12 mois. L'OMS (WHO 2010) considère qu'il s'agit d'un bon indicateur de la qualité globale des soins du VIH parce que :

- il témoigne à la fois de la qualité des soins médicaux et du bon suivi des

patient-e-s, mais aussi de l'accessibilité des services.
- les résultats négatifs (décès, perdus de vue) se produisent principalement dans la première année suivant le début du traitement (WHO 2012).

Le maintien sous traitement à 12 mois a été calculé en utilisant le registre informatique du programme ARV au Swaziland (APMR). C'est une base de données électronique comprenant les informations de base et le suivi de toutes les visites pour les patient-e-s sous ARV. Pour tous les patient-e-s compris dans la population de l'étude, les résultats à 12 mois après le début de leur traitement ont été encodés comme suit :

- Maintenu-e dans la thérapie ARV
- Décédé-e
- Transféré-e
- Traitement arrêté
- Perdu-e de vue

Les patient-e-s transféré-e-s (en provenance d'autres programmes ou partis vers d'autres programmes) ainsi que les patient-e-s dont le traitement a été arrêté sur décision médicale (échec thérapeutique non lié au type d'intervention) ont été exclus de notre analyse. Les patient-e-s ont été considérés comme maintenus dans la thérapie tant qu'ils et elles n'étaient ni décédé-e-s ni perdu-e-s de vue. Les patient-e-s étaient considéré-e-s comme perdu-e-s de vue si aucune visite les concernant n'avait été enregistrée dans l'APMR dans les 90 jours suivant la date de la dernière visite programmée.

Notre analyse était donc limitée aux patient-e-s qui avaient débuté leur traitement ARV en 2011 sur la même zone que pour l'étude de coût (service ARV du CSS de Nhlangano et cinq des sept CSP de la zone de santé). Les patient-e-s qui ont été transféré-e-s dans la zone d'étude en 2011 (102 personnes) ont été exclu-e-s de l'analyse ainsi que celles et ceux qui ont été transférés du CSS vers un CSP (13). Certaines exclusions ont également été faites en raison d'erreurs de données. La taille de l'échantillon final était de 1 140 patient-e-s.

Une analyse descriptive des caractéristiques des populations étudiées a été réalisée. Les comparaisons statistiques pour les facteurs potentiels de

confusion (tels que le sexe, le taux de CD4 et la tranche d'âge) entre les populations de patient-e-s du CSS et des CSP ont été faites en utilisant le test de Chi² de Pearson pour les variables catégorielles et les tests de Student ou de Kruskal-Wallis pour les variables quantitatives (selon que leur distribution est normale ou non). Une régression logistique inconditionnelle (en utilisant un résultat binaire : « maintenu dans le traitement à 12 mois » versus « non maintenu ») a été réalisée pour calculer les Odds ratios (rapport de chance) et les intervalles de confiance à 95 %. L'analyse statistique a été réalisée en utilisant Stata/SE (StataCorp, Texas, États-Unis, version 12.1.)

Principaux résultats

Notre étude montre que les soins de santé ARV dans les CSP sont à la fois plus efficaces (88 % maintenus dans la thérapie après 12 mois de traitement dans les CSP contre 76 % dans le CSS) et plus cher (311 $ par patient-e/année dans les CSP contre 233 $ dans le CSS). L'ICER s'élève à 652 $, ce qui représente le coût supplémentaire que le prestataire de services doit payer pour sauver une année de vie supplémentaire dans l'approche CSP, en comparaison de l'approche CSS. L'approche CSP domine donc l'approche CSS, car plus efficace, mais est-elle pour autant coût-efficace ?

La réponse à cette question est très subjective, car cela dépend du prix que l'on est prêt et/ou en mesure de payer pour sauver une année de vie supplémentaire. L'Organisation Mondiale de la Santé (OMS) a longtemps préconisé de comparer l'ICER avec le Produit National Brut (PNB) par habitant pour juger du coût-efficacité d'un programme (WHO, 2002). Cependant depuis 2015 cette préconisation, d'ailleurs fort controversée, a été abandonnée en faveur d'une démarche d'analyse plus fine du contexte du programme (Marseille 2015).

Dans le cadre de cette étude, MSF et le MOH ont considéré que le coût à payer pour sauver une année de vie supplémentaire était suffisamment bas pour pouvoir qualifier l'approche CSP de coût-efficace par rapport à l'approche CSS.

Analyse réflexive et conclusion

Le principal intérêt de cette évaluation économique rétrospective est qu'elle repose sur des données réelles d'un programme grandeur nature. Les résultats de cette évaluation économique ainsi que des trois autres volets de l'évaluation du programme de Shiselweni ont d'ailleurs fait l'objet :

- d'un rapport d'étude largement diffusé au sein de MSF
- d'une présentation orale lors de la 17^{e} conférence internationale sur le SIDA et les MST en Afrique (ICASA) en décembre 2013
- d'une présentation générale lors d'une conférence organisée par MSF au Swaziland en mai 2014 et à laquelle étaient conviés les représentant-e-s du MSP ainsi que des ONG internationales engagées dans la lutte contre le SIDA au Swaziland.

Le principal inconvénient de notre démarche est que l'évaluation n'avait pas été prévue dès le lancement du programme. Il a ainsi fallu faire avec les données collectées de façon routinière dans le programme, données initialement collectées pour un autre but et qui n'étaient donc pas forcément adaptées aux desseins de notre évaluation économique, ou qui ne répondaient pas forcément à nos besoins dans la forme sous laquelle elles ont été collectées.

Nous avons dû revoir le cadre de l'étude à la baisse par rapport aux ambitions initiales. L'objectif initial du commanditaire était de mesurer le coût-efficacité d'une approche décentralisée versus une approche centralisée. Il aurait fallu idéalement comparer deux zones de santé ayant adopté chacune l'une des deux approches. Ce n'était techniquement pas possible. Nous avons donc limité l'évaluation économique à une comparaison entre des patient-e-s sous ARV suivis dans un CSS ou dans un CSP. Pour des raisons de faisabilité et de contraintes de temps, nous avons dû également limiter l'évaluation au point de vue du prestataire de services, alors que la perspective du bénéficiaire aurait également été intéressante. Pour ce faire, il aurait fallu idéalement interroger un minimum de 200 patient-e-s sélectionnés de façon aléatoire sur tout le territoire afin d'obtenir des données fiables et statistiquement significatives. Nous pensons cependant que notre évaluation reste pertinente, car nous avons réussi à montrer que

l'approche CSP est coût-efficace malgré ces deux limites qui minimisent à notre avis la mesure de son efficacité.

Pour estimer le coût, nous avons suivi la méthode des ingrédients. Le choix des ingrédients que nous avons retenu est basé sur notre expérience d'estimation de coût par la méthode des coûts complets dans d'autres programmes SIDA dans des contextes similaires. Ce sont ces évaluations antérieures qui nous ont permis de distinguer les facteurs de coûts importants qu'il convient d'estimer avec soin de ceux moins importants que l'on peut négliger. Se baser uniquement sur la méthode des ingrédients sans avoir cette connaissance antérieure peut conduire à des biais d'estimation, à négliger des coûts importants, mais peu visibles (partie immergée de l'iceberg) ou au contraire à perdre un temps précieux à estimer des coûts négligeables (l'arbre qui cache la forêt).

Notons que l'accès à l'information financière n'est pas toujours aisé. Nous n'avons eu aucun mal à obtenir ces informations de la part de MSF qui était le commanditaire de l'étude et qui nous a donc ouvert tous ses livres de comptes en toute transparence. En revanche, le MSP ne nous a pas laissé accéder à ses données comptables. Comme le principal coût engendré par le MSP est celui du personnel de santé, nous avons fait nos estimations sur la base de la grille des salaires de la fonction publique au Swaziland qui relève de l'information publique.

Une autre difficulté concernant l'estimation des coûts est relative à l'abondance d'information et de sources d'information parfois contradictoires. Il convient de trianguler les données au moyen d'entrevues pour trier le bon grain de l'ivraie. L'utilisation d'Excel et des tableaux croisés dynamique nous permet également d'avoir une approche holistique de l'information du principal en zoomant sur certaines spécificités en cas de besoin sans pour autant nous noyer dans les détails.

Concernant la mesure de l'efficacité, l'indicateur que nous avons choisi est un indicateur standard de suivi des patient-e-s sous ARV. L'APMR, outil de suivi utilisé par le programme, est donc tout à fait adapté à son calcul. La difficulté réside ici ailleurs : comment savoir si l'efficacité supérieure d'une approche par rapport à l'autre dépend de l'approche en question ou bien d'autres facteurs de confusion ? Comme nous l'avons vu, nous avons pris en compte les principaux facteurs de confusion identifiés dans des programmes SIDA, à savoir le sexe, l'âge et le CD4 des patient-e-s et nous avons ajusté

nos résultats en fonction. Un autre facteur de confusion potentielle est le fait que le soutien de MSF était plus intense au niveau des CSP qu'au sein du CSS. Nous avons poussé un peu plus loin nos investigations pour déterminer si ce surcroit de personnel de santé dans les CSP pouvait avoir eu une influence sur leur meilleure efficacité, auquel cas nous aurions mesuré, non pas l'efficacité de l'approche CSP, mais celle du personnel supplémentaire MSF. Notre conclusion est que les CSP sont en sureffectif et qu'une baisse de personnel n'entraînerait pas de baisse de l'efficacité. Cela nous a conduits à réaliser des analyses de sensibilité avec l'hypothèse d'une baisse des coûts de personnel au sein des CSP, améliorant encore le coût-efficacité de cette approche.

En conclusion, l'une de principales leçons apprises de cette étude est qu'il est préférable de prévoir une évaluation économique dès le lancement d'un programme afin de collecter de manière prospective les données adaptées à sa réalisation. Forts des enseignements tirés de la conduite de cette évaluation, nous avons adapté notre méthode pour l'évaluation économique d'une nouvelle initiative pilote de ce même programme SIDA de la région de Shiselweni. La nouvelle initiative lancée par MSF s'appelle « *Early Access to* ARV *for All* » (accès précoce aux ARV pour tous). Il s'agit de mettre sous traitement ARV les patient-e-s aussitôt qu'ils et elles sont testés séropositifs. Rappelons que le protocole actuel en vigueur au Swaziland, comme dans la plupart des pays d'Afrique australe, veut que l'on attende que les défenses immunitaires mesurées par les CD4 se dégradent en dessous d'un certain seuil avant de mettre les patient-e-s séropositifs sous traitement. Un des avantages d'une mise sous ARV précoce réside dans le fait que l'on risque moins de perdre de vue un-e patient-e sous ARV qu'un-e patient-e séropositif n'ayant pas commencé son traitement. Un autre avantage est qu'un-e patient-e sous ARV risque moins de contaminer d'autres personnes qu'un-e patient-e séropositif sans traitement.

Pour mesurer le coût-efficacité de cette nouvelle initiative, nous avons repris la même méthode que celle décrite dans ce chapitre, mais cette fois, comme nous avons pu établir le devis de l'évaluation avant même que ne commence la mise en route de l'initiative évaluée, nous avons pu 1) apporter plus de soins dans la définition et la sélection des populations test et témoin 2) prendre des dispositions en amont pour faciliter la collecte des données et la mesure de l'efficacité.

Références clés

Conteh, L. (2004). Cost and unit cost calculations using step-down accounting. *Health Policy and Planning*, 19(2), 12735. https://doi.org/10.1093/heapol/czh015.

Cet article décrit en détail la méthode des coûts complets appliquée aux programmes de santé.

Drummond, M. F. (dir.). (2007). *Methods for the Economic Evaluation of Health Care Programmes* (3e éd.). Oxford : Oxford University Press.

Ce livre permet de se familiariser avec la théorie et la pratique de l'évaluation économique. Il décrit les méthodes de mesure des coûts, des résultats, de la collecte des données pendant les études cliniques et indique comment gérer l'incertitude dans l'évaluation économique.

Marseille, E., Larson, B., Kazi, D.S., Kahn, J. G. et Rosen, S. (2015). Thresholds for the cost–effectiveness of interventions: alternative approaches. *Bulletin of the World Health Organization*, 93(2), 11824. https://doi.org/10.2471/BLT.14.138206

Cet article explique les inconvénients majeurs de la méthode préconisée pendant plus de 10 ans par le projet CHOICE pour évaluer le coût-efficacité d'un programme (comparaison de l'ICER avec le PNB par habitant) et recommande des approches alternatives.

WHO-CHOICE. (s. d). Quantities and unit prices (cost inputs) http://www.who.int/choice/cost-effectiveness/inputs/en/

Le site internet du projet CHOICE présente les méthodes et les outils conçus par l'OMS en matière d'évaluations économique, dont la méthode des ingrédients.

Références

Bedelu, M., Ford, N., Hilderbrand, K. et Reuter, H. (2007). Implementing antiretroviral therapy in rural communities: the Lusikisiki model of decentralized HIV/AIDS care. *The Journal of Infectious Diseases*, 196(Suppl. 3), S464-S468. https://doi.org/10.1086/521114

Bemelmans, M., Van Den Akker, T., Ford, N., Philips, M., Zachariah, R., Harries, A., ... et Massaquoi, M. (2010). Providing universal access to antiretroviral therapy in Thyolo, Malawi through task shifting and decentralization of HIV/AIDS care: Task shifting to support universal access to ART. *Tropical Medicine et International Health*, 15(12), 1 413-1420. https://doi.org/10.1111/j.1365-3156.2010.02649.x

Chan, A. K., Mateyu, G., Jahn, A., Schouten, E., Arora, P., Mlotha, W., Kambanji, M. et van Lettow, M. (2010). Outcome assessment of decentralization of antiretroviral therapy provision in a rural district of Malawi using an integrated primary care model: assessment of decentralization of antiretroviral therapy provision. *Tropical Medicine et International Health*, 15, 90-97. https://doi.org/10.1111/j.1365-3156.2010.02503.x.

Cohen, R., Lynch, S., Bygrave, H., Eggers, E., Vlahakis, N., Hilderbrand, K., ... et Makakole, L. (2009). Antiretroviral treatment outcomes from a nurse-driven, community-supported HIV/AIDS treatment programme in rural Lesotho: Observational cohort assessment at two Years. *Journal of the International AIDS Society*, 12(1), 23. https://doi.org/10.1186/1758-2652-12-23.

Loubiere, S., Boyer, S., Protopopescu, C., Bonono, C. R., Abega, S., Spire, B. et Moatti, J.P. (2009). Decentralization of HIV care in Cameroon: Increased access to antiretroviral treatment and associated persistent barriers. *Health Policy*, 92(23), 165-173. https://doi.org/10.1016/j.healthpol.2009.03.006

Mulamba, D., Fullem, A., Hirschhorn, L., Allers, C., Oser, R. et Rau, B. (2010). *Decentralizing antiretroviral treatment services at primary health care facilities: A guide to expanding access and care in resource-limited, generalized HIV epidemic settings*. PUBL? https://aidsfree.usaid.gov/sites/default/files/decentralizing_art_services_primary_hc_facilities.pdf

Ridde, V. et Dagenais, C. (dir.). (2012). *Approches et pratiques en évaluation de programmes : nouvelle édition revue et augmentée*. Montréal : Les Presses de l'Université de Montréal.

Turashvili, M., Becher, H., Kerschberger, B., Jouquet, G., Browne, M., Haye, J. L., ... et Lassovski, M. (2014). Decentralisation of HIV/TB care inShiselweni region of Swaziland: Making a difference. *Titre de la revue*,

volume(numéro), p.-p.
http://fieldresearch.msf.org/msf/handle/10144/325599

Joint United Nations Programme on HIV/AIDS. (2010). *Swaziland Country Progress Report*.
http://data.unaids.org/pub/report/2010/swaziland_2010_country_progress_report_en.pdf.

United Nations Development Programme. (2009). *Overcoming barriers: Human mobility and development. Human development report*. Houndmills: Palgrave Macmillan.

World Health Organization. (2002). *The world health report 2002: Reducing risks, promoting healthy life*. Geneva : World Health Organization.
http://www.who.int/whr/2002/en/

World Health Organization. (2009). *Global tuberculosis control: epidemiology, strategy, financing*. Geneva : World Health Organization.

Organisation mondiale de la Santé. (2010). *Guide sur les indicateurs pour le suivi et le bilan de la riposte du secteur de la santé au VIH/sida*. Geneva : World Health Organization.
http://www.who.int/hiv/data/ua10_indicator_guide_fr.pdf

World Health Organization. (2012). *Retention in HIV programmes: Defining the challenges and identifying solutions*. Geneva : World Health Organization.
http://www.who.int/hiv/pub/meetingreports/retention_programmes/en/

Résumé / Abstract

L'évaluation économique décrite ici a été achevée en 2013 et fait partie d'une évaluation plus globale menée par quatre évaluateurs et évaluatrices indépendants à la demande de Médecins sans Frontière (MSF) dans un double but de documentation des expériences acquises et de plaidoyer. Le programme évalué a débuté en 2007 dans la région de Shiselweni au Swaziland en partenariat avec MSF et le ministère de la Santé publique. Ce programme de traitement du SIDA avait la particularité d'être décentralisé et intégré au niveau des centres de santé primaire. Il s'agissait d'une démarche innovante au Swaziland où le traitement du SIDA était alors centralisé au

niveau des centres de santé secondaire. Les approches décentralisées du traitement du SIDA dans les pays en développement ont fait l'objet de plusieurs études montrant leur efficacité. Le but de la présente évaluation était d'y ajouter un volet économique. Nous avons donc mené une étude rétrospective de coût-efficacité du point de vue du prestataire de service avec calcul du ratio de coût efficacité incrémental (ICER). Les principales difficultés pour une évaluation telle que celle-ci résident 1) dans l'accès aux données financières des prestataires de service 2) dans la mesure du gain d'efficacité liée à l'approche alternative en isolant les éventuels facteurs de confusion.

***

The economic evaluation described here was finalized in 2013 and is part of a more comprehensive evaluation conducted by four independent evaluators and sponsored by Médecins sans Frontières (MSF) for the purpose of documenting past experiences and advocacy. The evaluated program started in 2007 in the Shiselweni region of Swaziland in partnership with MSF and the Ministry of Public Health. This AIDS treatment program had the specificity of being decentralized and integrated at the level of primary health centers. This was an innovative approach in Swaziland where AIDS treatment was centralized at the secondary health center level. Decentralized approaches to AIDS treatment in developing countries have been the subject of several studies showing their effectiveness. The purpose of this evaluation was to add an economic component. We have therefore conducted a retrospective cost effectiveness study from the service provider perspective with calculation of the incremental cost-effectiveness ratio (ICER). The main challenges for an evaluation such as this one reside 1) in the access to the financial data of the service providers 2) to the measure of the effectiveness gain compared to the alternative approach and isolating possible confounding factors.

***

Guillaume Jouquet est consultant indépendant en économie de la santé et en management de projet. Il a commencé sa carrière dans l'humanitaire en 2005 avec Médecins sans frontière en tant que coordinateur adjoint d'un programme SIDA en Afrique du Sud. Depuis, il a accompagné et évalué quelques dizaines de projets humanitaires de santé principalement en Afrique subsaharienne.

Citation

Guillaume Jouquet. (2019). L'analyse coût-efficacité. Une intervention de décentralisation des soins VIH/SIDA à Shiselweni, Swaziland. In *Évaluation des interventions de santé mondiale. Méthodes avancées*. Sous la direction de Valéry Ridde et Christian Dagenais, pp. 311-332. Québec : Éditions science et bien commun et Marseille : IRD Éditions.

13. L'analyse spatiale

Un cas d'intervention communautaire de lutte contre le moustique Aedes aegypti au Burkina Faso

EMMANUEL BONNET, SAMIRATOU OUÉDRAOGO ET DIANE SARÉ

L'analyse spatiale

Définition de la méthode

Méthode d'analyse permettant de mesurer le degré de ressemblance d'un lieu avec ses voisins.

Forces de la méthode

Les résultats cartographiques peuvent être expliqués facilement à des décideuses et décideurs et au grand public

Défis de la méthode

- Elle ne permet pas de prendre en compte d'autres facteurs pouvant avoir un impact sur les résultats de l'intervention (ex. : caractéristiques du milieu, changement pluviométrique et/ou atmosphérique, etc.).
- Il est nécessaire d'intégrer dans le protocole d'évaluation l'acquisition de données géolocalisées à partir d'un échantillon géographique.
- Des compétences de base en analyse spatiale sont nécessaires.

Les méthodes d'analyse spatiale ont pour objectif de déterminer les caractéristiques de la distribution spatiale des individus géographiques (les ménages par exemple) ou de leurs valeurs (par exemple : le nombre de cas de dengue dans le ménage). Ces caractéristiques peuvent être analysées globalement, c'est-à-dire à une échelle importante comme celle d'une ville entière, pour laquelle les méthodes mettront en évidence le nuage de points des individus géographiques, les relations spatiales entre les valeurs et les individus (dépendance spatiale), et les caractéristiques géométriques (l'organisation des valeurs dans le territoire, comme un gradient nord-sud par exemple) du nuage de points ou des valeurs. Lorsque les valeurs étudiées sont quantitatives, il est possible d'analyser la distribution géographique des valeurs associées aux individus en cherchant notamment s'il existe une relation avec la proximité entre ces individus. L'analyse prend la forme d'un

indice qui exprime s'il existe ou non une corrélation entre la valeur des individus et leur proximité. Il permet par exemple de caractériser si les valeurs qui se ressemblent ont tendance à être proches dans le territoire ou, au contraire, sont totalement dispersées. Si les valeurs sont regroupées, on parle alors d'une autocorrélation spatiale positive. Si elles sont dispersées, on parle d'autocorrélation spatiale aléatoire. Cette méthode vise donc à caractériser la dépendance spatiale des valeurs.

De nombreuses situations, en particulier sanitaires, ne sont pas homogènes (on parle aussi de stationnarité). On pense notamment à la concentration de cas de maladie dans des localités très précises, quelques rues d'un quartier par exemple. Imaginons un quartier dont la marge est proche d'une étendue d'eau. Les cas de paludisme existent partout dans celui-ci, mais sont davantage présents et concentrés dans la rue qui borde l'étendue d'eau. La distribution spatiale du paludisme serait dans ce cas hétérogène à l'échelle du quartier. Cependant, les mesures globales de l'autocorrélation spatiale, à l'échelle du quartier dans son ensemble, peuvent occulter ce phénomène de concentration localisé lorsque l'autocorrélation varie d'un lieu à l'autre. L'indice d'autocorrélation spatiale obtenu pourrait signifier que la distribution spatiale des cas de paludisme est homogène dans le quartier, car le calcul, appliqué sur la globalité du territoire, pourrait minimiser la concentration identifiée dans la rue proche du barrage. On dit dans cette situation qu'il existe des associations spatiales de valeurs locales significatives qui ne sont pas détectables par la mesure classique de l'autocorrélation spatiale globale définie par Moran par exemple.

Un des objectifs de l'analyse spatiale est de détecter des formes ou des lieux singuliers où se produisent des concentrations locales de valeurs élevées, appelées aussi agrégats (d'individus ou de valeurs). Des méthodes permettent ainsi une analyse locale très précise et utile pour le ciblage de certaines actions par exemple. Luc Anselin (Anselin, 1995) a développé un indice pour pallier l'effet d'occultation des autocorrélations locales par les indices globaux : les indicateurs locaux d'association spatiale (LISA) (*Local indicator of spatial association*).

Les LISA permettent de mesurer s'il existe des lieux pour lesquels des valeurs associées aux ménages présentent une relation significative avec la proximité entre ces lieux et s'il existe parmi ces lieux des regroupements de valeurs très élevées ou très faibles. Ces indicateurs mesurent donc la concentration des valeurs numériques mesurées sur un ensemble de points.

Le calcul se base sur celui de l'autocorrélation globale de Moran, comme la moyenne des produits des valeurs normalisées des couples de points, pondérées par un poids spatial qui est représenté par la distance. L'analyse LISA s'effectue sur chaque ménage, représenté par un point, et sur les relations de chaque ménage avec ses voisins. Il consiste d'abord à calculer un indice pour chaque point (ménage) à partir de son voisinage. Les indicateurs statistiques expriment le degré de confiance et d'importance que l'on peut accorder aux informations acquises. Dans le cas d'unités spatiales irrégulières pour lesquelles seule la matrice de contigüité est prise en compte, le poids est égal à 1 lorsque les unités sont contiguës et 0 pour les autres. Ainsi, autour du point observé, le LISA local détermine l'étendue d'une région (*cluster*) formée par des points voisins qui possèdent des valeurs proches de la sienne. La somme des LISA locaux est proportionnelle à l'indicateur global. L'indicateur local se calcul de manière similaire à l'indice de Moran. L'interprétation des LISA obtenus se base sur les hypothèses statistiques de normalité. Il est donc nécessaire de tester la signification statistique des structures en cluster de la région. Une valeur du p est calculée pour chaque point : plus elle faible, plus l'éventuel *cluster* qu'elle forme avec ses voisins possède une existence statistiquement significative (Caloz et Collet 2011). Il en résulte cinq cas. Le premier, c'est lorsque la valeur p ne permet pas de rejeter l'hypothèse nulle : le point n'est pas considéré comme significatif pour être agrégé avec un autre point. Dans les quatre autres cas, les indices de LISA sont utilisés pour une typologie de quatre agrégats, en fonction de la valeur de l'individu et de la valeur du voisinage. Ces agrégats sont représentés dans le tableau 1, avec la couleur utilisée habituellement pour leur cartographie (Oliveau 2010).

		Valeur de l'individu statistique	
		Faible	Forte
Valeur moyenne du voisinage	Forte	Association spatiale négative	Association spatiale positive
	Faible	Association spatiale positive	Association spatiale négative

Tableau 1. Indices de LISA

Ce type de méthode est donc adapté, mais rarement utilisé, pour évaluer les effets d'une intervention dans l'espace, à des échelles géographiques très fines comme la localisation des ménages par exemple. Les valeurs analysées à l'échelle des ménages (coordonnées géographiques) peuvent représenter un nombre de personnes atteintes par une maladie par exemple (donnée quantitative). Les LISA seront calculés et représentés cartographiquement avant et après l'intervention afin d'évaluer les effets.

Analyse spatiale des effets d'une intervention communautaire de lutte contre la dengue à Ouagadougou

À la suite à différentes phases d'émergence de la dengue à Ouagadougou au Burkina Faso, il a été décidé de développer et de réaliser une intervention communautaire de lutte contre le moustique *Aedes aegypti*, principal vecteur de la dengue en milieu urbain.

Site d'étude et participant-e-s

L'étude a été réalisée dans deux quartiers comparables de Ouagadougou (Tampouy et Juvenat), choisis parmi cinq zones identifiées dans la ville. En effet, dans les deux quartiers, il existe des profils socioéconomiques fortement contrastés : des ménages aisés vivant dans des maisons en béton modernes totalement alimentées par les réseaux d'eau et d'électricité, mais aussi des ménages dont le niveau de vie est très bas dans des maisons construites à base d'argile parfois sans commodités de base (Observatoire des populations de Ouagadougou 2013). Ces quartiers sont aussi caractérisés par des zones où la gestion des déchets n'est pas régulièrement assurée et où les détritus s'accumulent dans certaines rues, favorisant ainsi la prolifération des moustiques.

Séance de causerie éducative dans un petit marché du quartier Tampouy (Crédit : Diane Saré)

Tampouy, qui se trouve dans la partie nord-ouest de la ville, a été choisie au hasard pour recevoir l'intervention, tandis que Juvenat, du côté est de la ville, a été choisi comme quartier « contrôle », pour en comparer les situations. Dans chaque quartier, nous avons délimité une zone d'un rayon d'un kilomètre autour du centre de santé primaire (considéré comme un point de repère). À Tampouy, l'intervention a été menée dans ce périmètre. Nous avons utilisé les coordonnées géographiques des ménages pour tirer au sort dans ce périmètre (tirage au sort sans remplacement) 287 ménages à Tampouy et 289 ménages à Juvenat, afin de collecter les données entomologiques (larves et nymphes de moustiques) nécessaires à l'évaluation de l'impact de l'intervention.

La théorie de l'intervention

1. La mobilisation et l'organisation : cette étape a permis d'établir le contact avec les leaders communautaires et les associations formelles et informelles œuvrant dans la zone de l'intervention à Tampouy pour les informer de la situation de la dengue dans leur quartier. L'étape a favorisé une prise de conscience des personnes influentes de la situation de cette maladie et leur implication dans la lutte.
2. La planification opérationnelle : elle consistait en un développement participatif de l'intervention impliquant les personnes influentes et les associations disposées à y participer.
3. L'action communautaire : elle consistait en la mise en œuvre des activités de l'intervention par les membres de la communauté (animateurs communautaires et troupe de théâtre) au profit de la communauté. Ces activités visaient à inciter une participation effective des ménages et de la population.
4. Le suivi/évaluation mené par la communauté : il a permis une meilleure implication de cette dernière.

L'interaction de ces différentes composantes a eu pour effets proximaux une augmentation des connaissances sur la dengue et les maladies fébriles à transmission vectorielle, un renforcement de la capacité d'action de la communauté de l'aire de l'intervention, et la réduction des indices entomologiques. Ce chapitre ne vise pas à décrire ces résultats[1].

Les effets distaux attendus sont le maintien de la lutte contre les moustiques, ce qui conduira à une réduction de la population des vecteurs de la dengue et du paludisme et ainsi à une réduction des épisodes et des hospitalisations liées à la dengue et au paludisme.

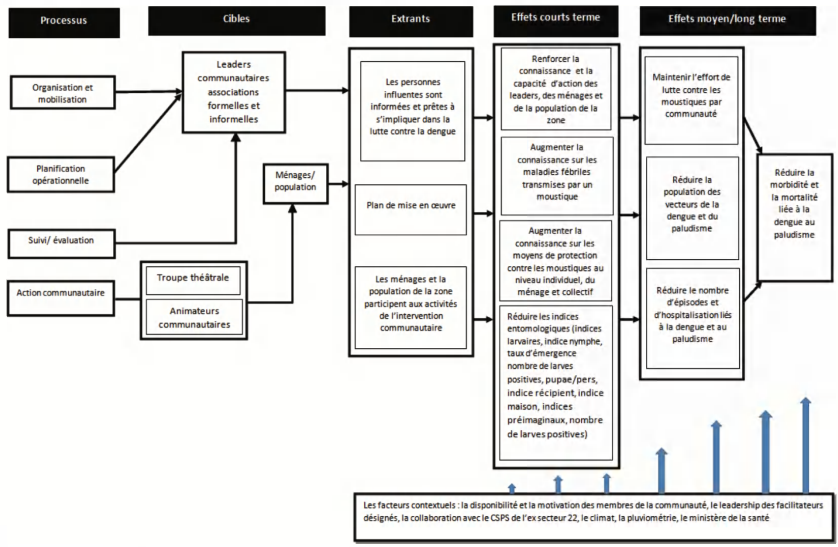


Figure 1. Modèle logique de l'intervention communautaire de lutte contre la dengue à Tampouy (ex secteur 22 de la ville de Ouagadougou)

Devis d'évaluation

Dans le tableau 2, nous résumons le devis d'évaluation de cette intervention dans le contexte spécifique de ce chapitre.

1. Pour plus d'information, voir le site http://www.equitesante.org/dengue-interventions-communautaires/

Tableau 2. Résumé du devis d'évaluation de l'intervention communautaire de lutte contre la dengue

Objectif d'évaluation	Questions d'évaluation	Planification
Évaluer l'impact de l'intervention sur la densité et la distribution géographique des larves et des nymphes d'*Aedes aegypti* dans le quartier de Tampouy	Est-ce qu'à la suite de l'intervention le nombre de larves et de nymphes a diminué à Tampouy par rapport à « Juvenat » qui n'a pas reçu d'intervention? Est-ce que suite à l'intervention, des zones de densité de larves et de nymphes persistent à Tampouy par rapport à « Juvenat » qui n'a pas reçu d'intervention?	Enquêtes entomologiques avant l'intervention, octobre 2015 : les larves et les nymphes d'*Aedes aegypti* ont été recherchées et collectées par les enquêteurs dans chaque concession visitée des deux quartiers. Juin à début octobre 2016 : réalisation des activités de l'intervention à Tampouy Enquêtes entomologiques après l'intervention fin octobre-début novembre 2016 : les larves et les nymphes d'*Aedes aegypti* ont été recherchées et collectées par les enquêteurs dans chaque concession visitée dans les deux quartiers.

Application de la méthode LISA

Les indicateurs locaux d'association spatiale permettent une exploration spatiale locale de l'impact de l'intervention communautaire de lutte contre l'*Aedes aegypti* au sein du quartier Tampouy de Ouagadougou. À partir des données des enquêtes entomologiques réalisées avant et après l'intervention, la procédure d'analyse se déroule en trois étapes : 1) la cartographie de la distribution spatiale du nombre de larves et de nymphes, 2) la mesure de cette distribution et 3) le calcul des indicateurs locaux d'association spatiale et leur cartographie.

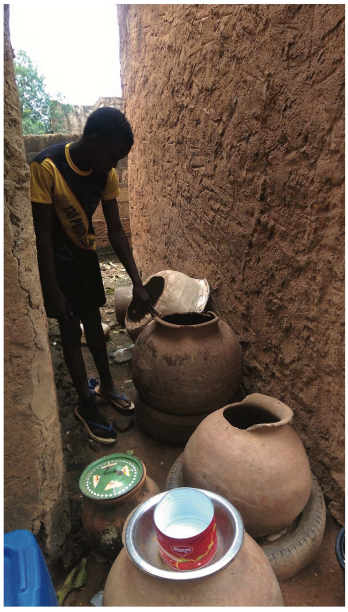
Étape 1 : Analyse de la distribution spatiale du nombre de larves et de nymphes par la cartographie

La première étape est de produire une cartographie des valeurs dans chacun des quartiers avant et après l'intervention en utilisant une échelle de

représentation commune. Ces représentations graphiques sont longuement exposées dans plusieurs manuels de cartographie qui permettent pour chaque type de variable et chaque type d'objet géographique (points, lignes, surfaces) de choisir la bonne représentation graphique (Bertin 1967; Bailly 1992).

Plusieurs logiciels, propriétaires (ArcGIs ou Mapinfo) ou libres (QGIs), permettent de réaliser la cartographie du nombre de larves et de nymphes en utilisant une taille de points proportionnels au nombre. Il est nécessaire de disposer des coordonnées géographiques de chaque ménage dans lequel a été réalisée l'enquête entomologique et pour lesquels un nombre total de larves et de nymphes a été calculé. À ce titre, pour éviter les biais d'interprétation, un plan de sondage spatial (tirage aléatoire dans l'espace) pour la collecte entomologique est nécessaire.

Les cartes (figure 1) représentent tous les ménages où les données ont été collectées. Si aucune larve n'a été relevée, le ménage est représenté par un point bleu sur la carte, il n'exprime aucune quantité. Pour les autres ménages, le point noir représenté est graphiquement proportionnel au nombre de larves et de nymphes calculé. La comparaison des cartes permet d'observer plusieurs phénomènes.



Identification des gîtes larvaires lors du porte-à-porte (Crédit : Diane Saré)

Lors de la phase initiale, avant l'intervention, on observe un nombre de larves et de nymphes important dans la zone d'intervention et la zone contrôle. Dans chacune d'elles, des zones apparaissent avec un nombre de larves et de nymphes plus important que dans d'autres parties des quartiers. Cela illustre que les larves d'*Aedes aegypti* sont présentes partout à Ouagadougou et qu'il existe, pour des raisons multiples, des « poches » dans lesquelles ces larves sont plus nombreuses.

Lors de l'enquête finale après l'intervention, on observe une diminution importante du nombre de ménages avec des gîtes larvaires et de nymphes à Tampouy, le secteur d'intervention. Dans le secteur contrôle, Juvenat, qui n'a pas reçu d'intervention, on observe que les gîtes identifiés lors de la phase initiale sont majoritairement présents, et l'apparition de nouveaux gîtes. Notons que les effets ne sont pas associés à une pluviométrie plus faible dans une zone par rapport à une autre, et d'une année à l'autre, et donc d'une plus faible productivité de gîtes larvaires.

Cette première étape est simple à réaliser, elle apporte une représentation cartographique du phénomène. Cependant, pour caractériser les « poches » et mesurer s'il existe une concentration du nombre de gîte et de larves dans certaines parties des quartiers et comment elle a évolué après l'intervention, il faut réaliser des mesures globales puis locales des valeurs associées aux ménages.

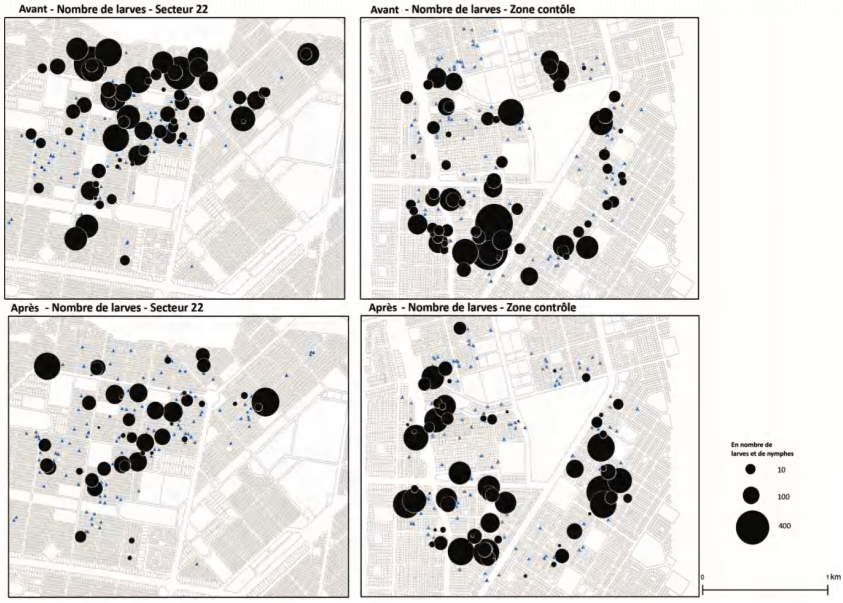


Figure 2. Cartographie du nombre de larves et de nymphes dans les quartiers de l'étude pré et post intervention

Étape 2 : Mesure de la distribution spatiale du nombre de larves et de nymphes au sein des quartiers de l'étude à Ouagadougou

Pour cette étape, on mesure l'autocorrélation spatiale pour déterminer s'il existe une corrélation entre la valeur des objets et les relations métriques ou topologiques entre ces objets. Pour rendre compte des valeurs voisines, on utilise des indices de corrélation, dont l'indice de Moran, le plus utilisé. Il se définit comme la moyenne des produits des valeurs normalisées des couples de points, pondérée par la distance entre les deux points (Moran, 1948).

$$I_{Moran} = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(x_i - \bar{x}\right)\left(x_j - \bar{x}\right)}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$$

où *i,j* = unité spatiale ; *n* = nombre d'unités spatiales ; x_i est la valeur de la variable dans l'unité *i* ; $\bar{x}$ est la moyenne de *x* ; et w_{ij} sont les éléments de la matrice d'interactions spatiales, définie sous la forme de la contigüité, les distances ou les frontières communes.

Cet indice mesure donc à quel point, pour l'ensemble de l'échantillon, la valeur des individus est proche ou éloignée de celle de leur voisinage par rapport à celle de la moyenne. Cette mesure est globale (Cliff, 1969) puisqu'elle porte sur la similarité des lieux avec leur voisinage pour l'ensemble de l'échantillon. Dans le cas de cette évaluation, l'indice de Moran indique si les habitations des ménages avec un nombre de larves et de nymphes important sont proches ou éloignées, révélant ainsi une autocorrélation spatiale positive ou négative. C'est la première étape avant de chercher à approfondir la localisation d'agrégats.

La valeur de l'indice de Moran dans le quartier de Tampouy avant l'intervention montre une autocorrélation spatiale positive (si l'indice est supérieur à zéro), relativement faible (i=0,10), mais avec une dispersion des valeurs de l'indice qui suggère que des agrégats sont possiblement présents. Après l'intervention, on remarque que l'autocorrélation spatiale est nulle (i=-0,02) ce qui signifie que les valeurs sont distribuées de manière aléatoire, sans autocorrélation, et démontre une évolution entre les deux phases de collecte et donc un effet de l'intervention. Dans le secteur contrôle, l'autocorrélation spatiale globale est également positive (i=0,05) avant l'intervention et nulle (i=-0,004) après. L'évolution globale dans les deux quartiers est donc similaire. Cependant, il est important de vérifier l'évolution au niveau local, c'est-à-dire au niveau de l'habitation des ménages, pour notamment vérifier s'il existe des agrégats, et si oui, comment ils ont évolué.

Étape 3 : Indicateurs locaux d'association spatiale du nombre de larves et de nymphes au sein des quartiers

Les indicateurs locaux d'association spatiale développés par Luc Anselin permettent de réaliser ces évaluations. Ils analysent la concentration des valeurs similaires et dissimilaires mesurées sur un ensemble de points ou calculées par agrégation spatiale.

$$I_i = \frac{\sum_j w_{ij} (p_i - \bar{p}) (p_j - \bar{p})}{\sum_i (p_i - \bar{p})^2}$$

où pi et pj sont les valeurs des unités spatiales i et j telles que i et j sont considérées comme voisines compte tenu de la mesure de leur degré de voisinage, pi est la valeur moyenne des unités spatiales et wij une mesure de la proximité des unités spatiales i et j.

La procédure peut se réaliser avec le logiciel libre Geoda (Anselin, 1995). Il permet la réalisation de toutes les étapes précédentes et ne nécessite pas le recours à un système d'informations géographiques pour préparer les données à utiliser. En revanche, les résultats sont exportables vers les SIG. Avant de faire l'analyse des indicateurs locaux d'association spatiale, il faut définir une matrice de poids (tableau qui recense toutes les distances entre tous les voisins) qui identifie les voisins. Les poids des voisins peuvent être construits de plusieurs manières en fonction du type d'objet géographique (un point ou une surface). Dans le cas de l'intervention de lutte contre la dengue présentée ici, les ménages sont représentés par des points. La matrice est donc calculée à partir d'une distance entre les points.

La dernière étape consiste à lancer l'analyse de LISA à partir de cette matrice de poids. Une statistique est calculée pour chaque unité, elle est basée sur la valeur des unités voisines. Il résulte deux types de cartes : une qui représente les valeurs de l'indice calculé et qui permet de visualiser comment l'indice varie dans le territoire, et une autre qui permet de mettre en évidence les relations statistiquement significatives des indices par une cartographie des agrégats qui illustre les similarités et dissimilarités de valeurs.

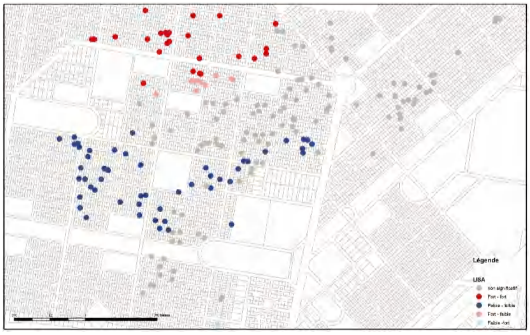


Figure 3a. Tampouy avant l'intervention

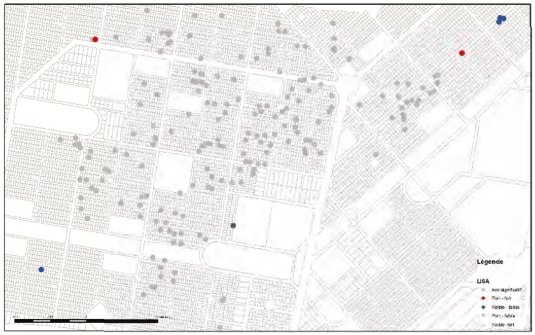


Figure 3b. Tampouy après l'intervention

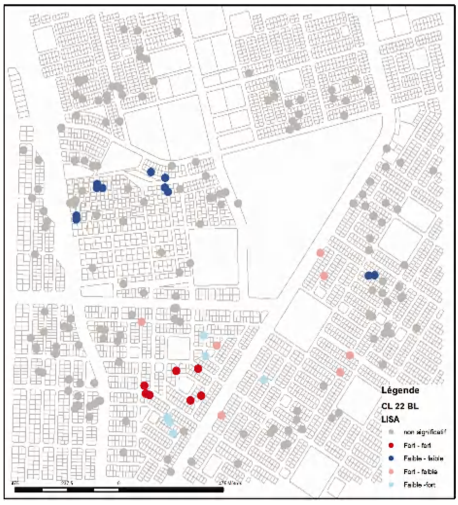


Figure 3c. Juvénat avant l'intervention

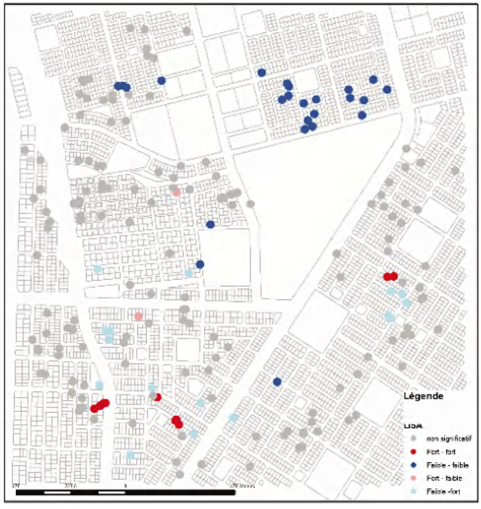


Figure 3d. Juvénat après l'intervention

Les cartes LISA (figures 3a, 3b, 3c et 3d) des valeurs de larves et de nymphes réalisées pour le quartier d'intervention Tampouy (en haut)

montrent qu'avant l'intervention, des agrégats de valeurs négatives (en bleu) et de valeurs positives (en rouge) étaient présents. Cela signifie qu'il existait des concentrations de deux types, très tranchées dans le quartier. En bleu, on voit une concentration de ménages qui avaient des gîtes larvaires et de nymphes relativement peu productifs, car ils avaient un faible nombre de larves et de nymphes. Cette concentration de valeurs en faible-faible signifie également que les ménages voisins avaient des caractéristiques similaires (faible nombre de larves). À l'inverse, au nord de la zone, la concentration de valeurs fortes est représentée par les points classés en forte-forte (valeurs fortes du ménage et de ses voisins). L'analyse LISA met donc en évidence une double concentration de valeurs de larves et de nymphes opposées. Cela ne signifie pas une présence ou une absence de larves et de nymphes, mais qu'il y avait des zones graduées de productions de larves et de nymphes dans l'aire étudiée avant l'intervention.

Après l'intervention, on constate que les analyses LISA révèlent peu d'agrégats de tous types et la disparation d'agrégats qui existaient avant l'intervention. L'analyse met donc en évidence que l'intervention a permis la réduction du nombre de zones de concentration de valeurs fortes et faibles de larves et de nymphes. L'analyse globale le confirme avec une autocorrélation spatiale nulle qui signifie que la présence de larves et de nymphes sur la zone d'intervention est désormais aléatoire. Cette analyse complète les analyses des épidémiologistes qui ont observé une réduction importante des indices entomologiques à Tampouy après l'intervention.

Dans la zone contrôle, on observe des agrégats qui persistent dans les mêmes lieux lors de la collecte après l'intervention et confirment l'augmentation de la production de larves et de nymphes identifiée par les autres analyses épidémiologiques.

Analyse réflexive et conclusion

L'évaluation de la dimension spatiale dans les interventions en santé publique est souvent limitée à une analyse des effets moyens de l'intervention qui varient en fonction des entités géographiques et des mesures approximatives de la distance. Les méthodes traditionnelles utilisées pour l'évaluation ne permettent pas de rendre compte de la variabilité spatiale des lieux et des effets des interventions. La méthode

utilisée dans ce chapitre est un exemple d'analyse spatiale qu'il est possible d'utiliser. Ces méthodes sont rarement mobilisées pour l'évaluation des interventions et davantage pour explorer la diffusion ou la concentration d'une maladie dans l'espace. Or, cet exemple démontre que les phénomènes peuvent être observés à différentes échelles et que les résultats globaux peuvent cacher des situations locales différentes. Ces analyses viennent compléter, par leur dimension spatiale, les autres résultats dans le cadre de l'évaluation d'impact d'une intervention. Généraliser l'utilisation des analyses spatiales dans l'évaluation des interventions est donc à recommander pour appréhender tous les effets à toutes les échelles.

Dans l'exemple présenté, la procédure utilisée s'est avérée efficace et n'a pas posé de difficultés dans la mesure où l'échantillonnage était géographique. C'est un des points fondamentaux des méthodes d'analyse spatiale. La qualité des résultats dépend d'un plan d'échantillonnage géographique pour assurer une interprétation correcte des résultats. L'un des enjeux de cette méthode est de retrouver les ménages lors de la phase post-intervention. Dans les pays où l'adressage des rues est inexistant, il faut avoir recours à des GPS et s'assurer que la géolocalisation est correcte et bien enregistrée. Dans les milieux urbains denses, certaines perturbations du signal brouillent les positions enregistrées dans les GPS et complexifient la possibilité de retrouver les ménages enquêtés en phase de pré-intervention. Lors de l'enquête auprès des populations il faut s'assurer d'enregistrer le numéro de téléphone de la personne enquêtée afin de servir d'alternative pour retrouver le ménage. Ne pas retrouver le ménage de l'enquête initiale est la principale faiblesse de la méthode sur le plan de sa mise en œuvre dans les pays à plus bas revenu. Les autres faiblesses sont relatives à la méthode en elle-même qui ne permet d'analyser qu'une variable à la fois, limitant ainsi la prise en compte d'autres variables pouvant être considérées comme des facteurs de confusion. Toutefois, il existe d'autres méthodes géographiques qui combinent plusieurs variables et seraient complémentaires (régressions géographiquement pondérées GWR) aux évaluations statistiques.

L'accès à ces méthodes est aujourd'hui facilité par la disponibilité de logiciels de Systèmes d'Informations Géographiques libres comme QGIS. Ils permettent de cartographier la géolocalisation des ménages d'une enquête, mais aussi de réaliser un échantillonnage géographique. Associés aux outils

libres « GéoDa » développés par Luc Anselin, la réalisation des LISA est également facilitée.

Références

Anselin, L., (1995). Local indicators of spatial association – LISA. *Geographical Analysis*, *27*(2), 93-115.

Bailly A. S. (1999). Pumain D. et Saint-Julien T., L'analyse spatiale T. 1. *Annales de Géographie*, (605), 105.

Bertin, J. (1967). *La sémiologie graphique : les diagrammes, les réseaux, les cartes*. La Haye ou Paris : Mouton ou Gauthier-Villars.

Caloz, R. et Collet, C. (2011). *Analyse spatiale de l'information géographique* (1e éd.). Lausanne : Presses polytechniques et universitaires romandes

Cliff, A. D. et Ord, K. J. (1969). The problem of spatial autocorrelation. Dans A. J. Scott (dir.), *Papers in Regional Science, Studies in Regional Science* (vol. 1, p. 25-55). Londres: Pion.

Moran, P. A. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society: Series B*, 10(2), 243-251.

Observatoire des populations de Ouagadougou, Projet Wellcome Trust 2008-2013. (2013). *Rapport scientifique sur la caractérisation sociologique des quartiers de l'OPO*. Repéré à http://www.issp02.issp.bf/opo/Publications/Rapport-scientifique-juin2013-Caracterisation-des-quartiers-Axe-qualite.pdf

Oliveau, S. (2010). Autocorrélation spatiale : leçons du changement d'échelle. *L'Espace Géographique*, 39(1), 51-64.

Résumé / Abstract / Resumen

Ce chapitre présente une analyse spatiale d'une intervention qui vise à évaluer si des valeurs similaires d'un résultat sont plus susceptibles de former des grappes dans une zone d'étude. Cette méthode appelée « indicateurs locaux d'association spatiale » (local indicators of spatial association – LISA) permet d'identifier les zones où les valeurs sont spatialement dépendantes les unes des autres. LISA a été utilisé pour évaluer

l'effet d'une intervention communautaire visant à contrôler le moustique responsable de la propagation de la dengue Aedes aegypti dans deux quartiers de Ouagadougou, capitale du Burkina Faso. L'évaluation finale de l'intervention à l'aide de cette méthode a montré que les groupes de larves ont disparu dans le quartier qui a reçu l'intervention alors qu'ils persistaient dans le quartier témoin.

***

This chapter present a spatial analysis of an intervention which aims at assessing whether similar values of an outcome are more likely to form clusters in a study area. This method called local indicators of spatial association (LISA) enable to identify areas where values are spatialy dependant with one another. LISA has been used to evaluate the effect of a community-based intervention which aims at controlling the mosquito responsable for spreading Dengue, *Aedes aegypti* in two neighborhoods of Ouagadougou the capital city of Burkina Faso. The final evalutation of the intervention using this method showed that larva clusters disappeared in the neighborhood which received the intervention while they persisted in the control neighborhood.

***

Este capítulo presenta un análisis espacial de una intervención para evaluar si los valores similares de un resultado tienen más probabilidades de formar grupos en un área de estudio. Este método, denominado « indicadores locales de asociación espacial » (LISA), identifica áreas en las que los valores dependen espacialmente unos de otros. LISA se utilizó para evaluar el impacto de una intervención comunitaria para controlar al mosquito responsable de la propagación del dengue Aedes aegypti en dos barrios de Uagadugú, la capital de Burkina Faso. La evaluación final de la intervención utilizando este método mostró que los grupos de larvas desaparecieron en el área que recibió la intervención mientras persistían en el área de control.

***

Emmanuel Bonnet est géographe de la santé et chargé de recherches à l'IRD (Institut de Recherche pour le Développement). Il a mené de nombreuses recherches en Afrique de l'Ouest et possède une expérience internationale dans les domaines de la surveillance des épidémies, des

technologies géomatiques, des systèmes d'informations géographiques et de l'analyse spatiale.

Samiratou Ouédraogo est épidémiologiste et chercheure en santé publique avec une expérience internationale en évaluation. Elle travaille sur de nombreux projets de recherche au Canada et en Afrique.

Diane Saré est médecin de santé publique. Elle possède une expérience professionnelle clinique tant en milieu communautaire qu'en milieu hospitalier. Elle possède également une expérience en mise en œuvre et en évaluation d'intervention communautaire.

Citation

Emmanuel Bonnet, Samiratou Ouédraogo et Diane Saré. (2019). L'analyse spatiale. Un cas d'intervention communautaire de lutte contre le moustique Aedes aegypti au Burkina Faso. In *Évaluation des interventions de santé mondiale. Méthodes avancées*. Sous la direction de Valéry Ridde et Christian Dagenais, pp. 333-351. Québec : Éditions science et bien commun et Marseille : IRD Éditions.

PARTIE V

L'ÉVALUATION DES PROCESSUS ET DE LA FIDÉLITÉ D'IMPLANTATION

14. L'analyse des processus de mise en œuvre

Une intervention complexe au Burkina Faso : le financement basé sur les résultats

VALÉRY RIDDE ET ANNE-MARIE TURCOTTE-TREMBLAY

L'analyse des processus de mise en œuvre

Définition de la méthode

L'analyse de la mise en œuvre vise à comprendre comment un ou plusieurs processus organisés dans un contexte particulier participent aux changements souhaités par une intervention au regard des moyens déployés.

Forces de la méthode

- Comprendre comment une intervention fonctionne dans son contexte réel et comment elle évolue dans le temps
- Évite l'erreur de type III, soit l'évaluation d'une intervention alors qu'elle n'a pas été mise en œuvre tel que prévu
- Utilise des devis rigoureux, comme des études de cas multiples longitudinales combinant plusieurs types de collectes de données complémentaires, tels que l'observation, les entrevues formelles et informelles ou la documentation

Défis de la méthode

- La quantité parfois importante de données qualitatives, l'arbitrage entre le renforcement de la validité interne *versus* externe, le temps de l'analyse et la complémentarité du qualitatif avec le quantitatif
- L'utilisation de modèles logiques (théorie de l'intervention) et de cadres conceptuels pour guider la collecte et l'analyse des données, mais sans s'y enfermer
- Innover et rester ouvert aux données empiriques émergentes

« Nous savons ce qu'il faut faire, mais nous ne savons pas comment le faire » : voici une phrase qui semble se retrouver de manière récurrente dans les rapports des acteurs et actrices de la santé mondiale depuis longtemps. En 2010 par exemple, les Nations Unies affirmaient que pour s'occuper de la

santé des femmes et des enfants « nous savons ce qui fonctionne[1] » (United Nations Secretary-General 2010). Pourtant, si le contenu des interventions théoriquement efficaces est relativement bien connu, leur niveau de couverture reste faible (Peters *et al.* 2009). De plus, les conditions de leur mise en œuvre sont très peu appréhendées. Une méta-analyse montre que l'efficacité théorique des interventions est réduite de 50 % à cause de multiples facteurs contextuels agissant sur la mise en œuvre (J. A. Durlak et DuPre 2008). Ainsi, il ne suffit pas de savoir si une intervention de santé mondiale est efficace dans un environnement contrôlé (efficacité théorique : *efficacy*), il faut aussi comprendre pourquoi elle fonctionne, comment, pour qui et dans quels contextes réels (efficacité pratique : *effectiveness*). C'est ici que la science de la mise en œuvre (*implementation science*) est d'un secours indéniable. Elle est définie comme un champ de recherche visant à comprendre les « facteurs associés à une mise en œuvre efficace » (Franks et Schroeder 2013). L'efficacité de la mise en œuvre nous renvoie à la capacité de l'intervention à organiser les activités essentielles à l'atteinte de ses objectifs, mais aussi, et peut-être surtout, à sa capacité à s'adapter aux acteurs et actrices et aux contextes tout au long de son histoire (Ridde, Delormier et Gaudreau 2007).

Pour reprendre les termes d'un auteur important de ce champ : « étudier la mise en œuvre des interventions n'est pas facile, mais est essentiel » (Durlak 2015b). Dans une revue des écrits sur le concept, Pfadenhauer et ses collègues (2015) définissent la mise en œuvre comme « un effort initié de manière délibérée et activement planifié dans l'intention de mettre en pratique un objet spécifique ». Dans ce livre sur les méthodes d'évaluation des interventions de santé mondiale et pour ce chapitre, la mise en œuvre est comprise comme un ou plusieurs processus organisés dans un contexte particulier afin de participer aux changements souhaités par une intervention (au sens large du terme : politique, programme, projet) au regard de ses moyens déployés. La mise en œuvre peut se distinguer de l'implantation, que certains considèrent comme un anglicisme, dans le sens où cette dernière est plus normative ou plus « simple » alors que la première est véritablement complexe et dynamique. L'évaluation de l'implantation sera utile pour vérifier que les activités ont été organisées tel que prévu (analyse

1. Toutes les traductions des citations en anglais sont des traductions libres des auteurs de ce chapitre

de la conformité) tandis que l'évaluation des processus permettra d'en comprendre la mise en œuvre, sa complexité, le rôle des acteurs et actrices sociaux et du contexte.

La science de la mise en œuvre vise donc à mobiliser des théories, concepts et méthodes pour mieux « comprendre quoi, qui et comment une intervention fonctionne dans un contexte réel » (Peters *et al.* 2013). Il n'existe évidemment pas de « contexte irréel », mais nous souhaitons ici plutôt faire référence au contexte naturel de mise en œuvre d'une intervention (Petticrew *et al.* 2005) comparativement à des contextes où les chercheuses et chercheurs et les intervenant-e-s tentent de contrôler les situations (le contexte et l'action) pour en prouver l'efficacité, usant souvent d'une approche expérimentale et de devis tels que les essais contrôlés randomisés.

On s'éloigne donc dans ce chapitre du concept de « *Implementation Research* », souvent impossible à traduire[2] et qui concerne surtout l'analyse de la manière dont les interventions tiennent compte et intègrent des données probantes dans la formulation de leur contenu (Nilsen *et al.* 2013). La revue *Implementation Science* en est l'étendard, même si, en utilisant ce titre, elle apporte une confusion supplémentaire dans un domaine qui n'en avait pas besoin. De plus, ce champ est aussi proche de celui du transfert des connaissances, lui aussi très peu développé dans les pays du Sud (Siron, Dagenais et Ridde 2015). Cependant, nous sommes d'accord avec le fait que « les recherches dans ces deux champs s'intéressent aux défis de concrétiser des intentions en changements souhaités » (Nilsen *et al.* 2013).

Dans ce chapitre, nous nous intéressons donc aux processus de mise en œuvre des interventions. Cependant, les recherches sur les processus de mise en œuvre en santé mondiale doivent mieux exploiter les théories, les cadres conceptuels et les approches des sciences sociales (Van Belle, van de Pas et Marchal 2017). Par exemple, Durlak propose une synthèse intéressante en suggérant huit domaines d'étude qui reprennent les fameux « *implementation outcomes* » de Proctor *et al.* (2011) : acceptabilité, adoption, pertinence, faisabilité, fidélité, coût, couverture, pérennité. Il suggère aussi

2. Tout comme le terme « implementation » qui est traduit de manière très différente en français par implantation, implémentation (Suisse, Belgique), mise en œuvre, etc. Dans ce livre, le terme anglais « implementation » est traduit par « implantation » quand on évoque une analyse de la conformité ou de la fidélité (voir le chapitre de Perez *et al.*) et « processus de mise en œuvre » dans le présent chapitre quand il s'agit de comprendre le fonctionnement de l'intervention.

23 facteurs influençant la mise en œuvre et 14 étapes dont il faut s'inspirer pour mettre en œuvre une intervention (Joseph A. Durlak 2015a; Figure 1). Si les approches inductives et empiriques sont essentielles pour faire émerger des discours et repérer des pratiques efficaces, le recours aux développements théoriques et conceptuels l'est tout autant (Olivier de Sardan et Ridde 2015), car il permet une approche plus systématique et plus globale ainsi qu'une meilleure organisation des données pour en faciliter l'analyse et la présentation.

Rappelons que parmi les études de promotion de la santé qui analysent les politiques publiques, seules 18 % font référence à un cadre théorique de ce champ (Breton et De Leeuw 2011). Des travaux ont montré, par exemple, que la théorie des courants de Kingdon (Ridde 2009), celle des acteurs et actrices de première ligne de Lipsky (Walt *et al.* 2008) ou encore celle de l'innovation de Rogers (Perez *et al.* 2013), toutes développées aux États-Unis, peuvent être adaptées au contexte de certains pays du Sud pour comprendre la mise en œuvre. En outre, le rôle des idées dans la mise en œuvre (contrairement à l'émergence) des interventions dans les pays du Sud n'a encore que très peu été abordé (Béland et Ridde 2016), alors que les écrits à cet égard dans les pays du Nord sont légion (Cox 2001).

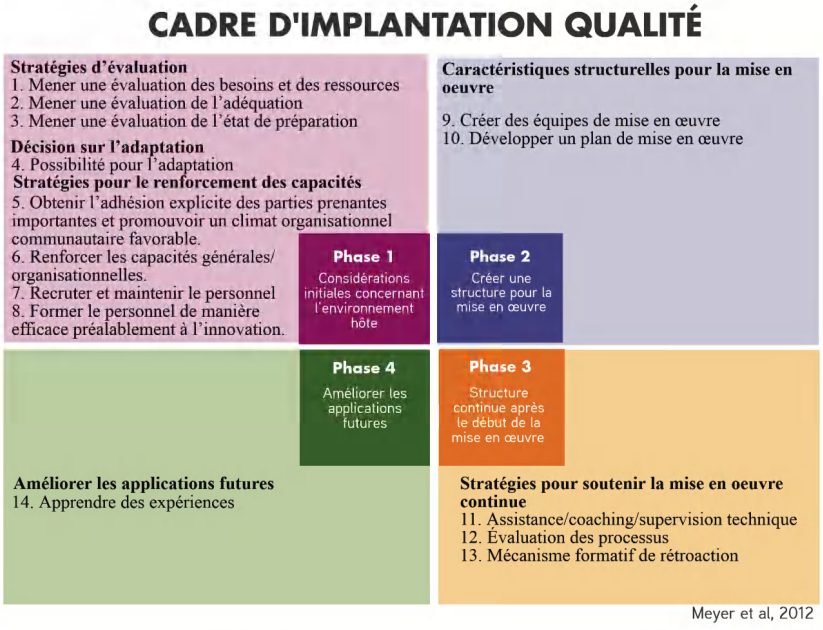


Figure 1. 14 étapes pour réussir la mise en œuvre selon Joseph A. Durlak (2015a)

La science de la mise en œuvre ne peut évidemment pas se développer seule, sans considérer les effets des interventions (ou les principes d'efficacité, pour reprendre les termes de Patton [2010]), au risque de tomber dans l'erreur de type trois, où l'on évalue une intervention alors qu'elle n'a pas été mise en œuvre tel que prévu (Dobson et Cook 1980). Il est important, par exemple, de mieux décrire le contenu des interventions ainsi que de rendre compte de la (classique pour les épidémiologistes) fidélité et de l'intensité de l'implantation (voir le chapitre de Perez *et al.*).

Ainsi, il est temps de prendre au sérieux la demande formulée il y a 10 ans en ce qui concerne les pays du Sud : « plus de recherches sur la mise en œuvre et spécialement les défis de la mise en œuvre des interventions qui visent plus d'équité » (Gilson et Raphaely 2008). C'est donc dans ce contexte que nous avons réalisé une analyse de la mise en œuvre d'une intervention complexe au Burkina Faso que nous décrivons dans la prochaine section.

Contexte et programme évalué

Depuis plusieurs années en Afrique, de nombreux pays se sont lancés dans des interventions de financement basé sur les résultats (FBR) pour améliorer la performance des systèmes de santé. Il existe de nombreux modèles de FBR, mais dans ce chapitre nous décrivons celui promu par la Banque Mondiale au Burkina Faso. Dans ce contexte, il s'agit de lier le financement des formations sanitaires et le paiement de primes au personnel de la santé à l'atteinte d'indicateurs de quantité et, souvent, de qualité des soins, préalablement définis. Ces indicateurs sont l'objet d'un processus de contrôle par des acteurs et actrices externes à la formation sanitaire. Les preuves scientifiques de l'efficacité ou de l'efficience du FBR restent limitées au Nord (Milstein et Schreyoegg 2016) comme au Sud (Turcotte-Tremblay *et al.* 2016 ; Witter *et al.* 2012). Les recherches sur sa mise en œuvre sont encore très rares en Afrique de l'Ouest francophone (Paul, Sossouhounto, et Eclou 2014). Elles ont des méthodes et des résultats très variables, mais confirment toutes le manque de connaissances à l'égard de la mise en œuvre du FBR en Afrique.

Au Burkina Faso, après un projet pré-pilote dans trois districts, le ministère de la Santé et la Banque Mondiale ont déployé une intervention FBR dans 12 districts supplémentaires en janvier 2014. Le Burkina Faso a

décidé d'innover en y associant des stratégies destinées à améliorer l'équité et soutenir l'accès aux soins des plus pauvres. Ainsi, dans 10 districts, un processus de sélection communautaire des indigent-e-s leur permettant d'être exemptés du paiement des soins a été organisé dans certains centres de santé sélectionnés de façon aléatoire. Deux autres districts organisent une assurance maladie à base communautaire associée à l'exemption du paiement des indigents. Tou-te-s les patient-e-s continuent de payer les soins au point de service sauf les indigent-e-s qui en sont exempté-e-s ou les femmes enceintes dont les accouchements sont subventionnés (80 %) par l'État depuis 2006 et gratuits depuis 2016. Rédigé avec le soutien de consultants de la Banque Mondiale, un guide de mise en œuvre de l'intervention de plus de 300 pages est disponible.

Une unité centrale appelée le Service Technique FBR (ST-FBR) a été créée pour appuyer la mise en œuvre du FBR à travers l'élaboration et le suivi des indicateurs, la supervision des vérifications, l'organisation d'ateliers de formation et l'analyse de données de performance. Des agences de contractualisation et de vérification (ACV) ont été mises en place pour vérifier les résultats quantitatifs des formations sanitaires, accompagner les agent-e-s de santé dans l'utilisation des outils du FBR, saisir les données et soumettre les factures des formations sanitaires pour validation. Les ACV effectuent les vérifications quantité mensuellement pour compter le nombre de prestations ciblées effectuées (~23 indicateurs). Les vérifications qualité sont effectuées trimestriellement sur plusieurs dimensions de la qualité des soins telles que les conditions d'accueil des patient-e-s, la disponibilité du matériel, la tenue des supports, la gestion des médicaments et la gestion financière (~104 indicateurs). Cette dernière requiert la collaboration des membres des comités de gestion (COGES) des formations sanitaires en raison de leur implication dans la trésorerie du Centre de Santé et de Promotion Sociale (CSPS). Divers outils (par exemple, un plan d'amélioration de la performance, outil d'indice) sont mis en place pour promouvoir la planification des activités et des finances. Les indicateurs de quantité sont achetés à des prix unitaires fixes et ceux pour la qualité le sont en fonction d'une cible minimale à atteindre.

L'organisation du FBR a été conçue pour faciliter son évaluation d'impact financée par la Banque mondiale selon un essai contrôlé randomisé des structures de première ligne, à savoir les CSPS. Quatre catégories de randomisation, comme autant de bras d'intervention, sont définies : (i) FBR 1 :

le centre de santé est payé pour les indicateurs d'activités atteints sur la base de prix fixés, mais qui peuvent fluctuer au cours du projet ; (ii) FBR 2 : le FBR 1 couplé à une intervention de sélection communautaire des indigent-e-s exempté-e-s du paiement des soins dont l'achat est effectué à un prix modéré ; (iii) FBR 3 : le FBR 2 à un prix d'achat plus élevé pour les indigent-e-s visant à encourager les initiatives innovantes pour augmenter leur utilisation des services ; (iv) FBR 4 : FBR 1 associé à une assurance maladie à base communautaire couplée à une intervention de sélection communautaire des indigent-e-s. Cette dernière composante du FBR est mise en œuvre uniquement dans une seule région sanitaire.

Pour les quatre modalités, un « bonus équité » a été mis en place entre les districts et les centres de santé. Ainsi, les prix d'achats varient sensiblement pour les centres de santé au sein de ces quatre catégories. Ces prix ont été calculés afin de notamment tenir compte de l'éloignement des centres de santé du chef-lieu du district et les besoins en ressources. Par exemple, une consultation curative d'une personne de plus de cinq ans peut être achetée entre 100 et 180 francs CFA[3] (100 francs CFA = 0,15 euros) selon les CSPS dans un même district.

L'analyse de la fidélité de l'implantation du FBR entreprise 12 mois après son démarrage montre que, globalement, les éléments importants et essentiels de la théorie d'intervention ont été implantés (Belaid *et al.* 2016). Cependant, nous avons noté des déficits d'implantation et des retards au niveau de la vérification de la performance et des paiements des subsides FBR, compromettant quelque peu l'efficacité théorique de la logique incitative. Bien que cette analyse de la fidélité soit essentielle, elle n'est pas suffisante pour comprendre le fonctionnement de l'intervention dans toutes ses subtilités. Ainsi, nous avons entrepris une analyse qualitative de la mise en œuvre du FBR au Burkina Faso selon un devis explicité dans la prochaine section.

Devis d'évaluation

Cette étude s'inscrit dans le courant des recherches interventionnelles

3. Communauté Financière Africaine.

et de l'évaluation fondées sur la théorie de l'intervention (Weiss 1998). La théorie de l'intervention FBR a guidé notre collecte et nos analyses de données. Au plan méthodologique, nous nous inscrivons dans la troisième génération de recherches sur la mise en œuvre des interventions en usant notamment d'études de cas multiples longitudinales (Saetren 2014). Cependant, au moment de la rédaction de ce chapitre, seulement la première de trois phases de collecte de données prévues pour les études de cas a été effectuée et est l'objet de cette présentation.

L'approche méthodologique est celle d'une étude de cas multiples (les CSPS) et contrastés avec plusieurs niveaux d'analyse imbriqués (les groupes d'acteurs et actrices concernés) (Yin et Ridde 2012). La sélection des cas a été explicitée ailleurs (Zombré *et al.* 2016). La recherche s'est déroulée dans trois districts sanitaires différents afin de représenter la diversité des processus de mise en œuvre du FBR. La collecte s'est déroulée dans un hôpital régional, deux hôpitaux de districts et 18 formations sanitaires de première ligne (CSPS). L'analyse pour les hôpitaux fait l'objet d'un autre rapport, car le contexte est différent. Une discussion de groupe avec les membres des trois équipes-cadres de district (ECD), soutenue par une analyse statistique des données de routine, a permis de classer les CSPS selon leur niveau de performance préalable au démarrage du FBR. La performance était ici comprise au regard des indicateurs d'utilisation des services disponibles dans ces données de routine dont la qualité a été vérifiée. Dans les deux districts qui organisent les FBR1, FBR2 et FBR3, un CSPS performant et un non performant ont été sélectionnés (n= 3 x 2 x 2). Dans le district organisant le FBR 1 et le FBR4, un CSPS performant, un CSPS moyennement performant et un CSPS non performant ont été sélectionnés (n= 3 x 2) (Figure 2). Dans chacun des districts, la recherche était sous la responsabilité d'un chercheur expérimenté du pays, socio-anthropologue, qui coordonnait la collecte des données réalisée par quatre assistant-e-s de recherche, responsables de la collecte des données dans deux ou trois CSPS du même type de FBR, sauf pour les hôpitaux où un seul assistant a réalisé la collecte de données. Ils et elles ont tous vécu 10 à 15 jours dans chaque village/ville où les formations sanitaires sont situées. Les trois collectes de données ont débuté à l'automne 2014, 2015 et 2016, respectivement. Un atelier de trois jours a réuni tous les chercheuses et chercheurs et assistants au démarrage de la recherche afin de coordonner la démarche, avec quelques séances de discussion avec des décideuses et décideurs et organisations mettant en œuvre l'intervention.

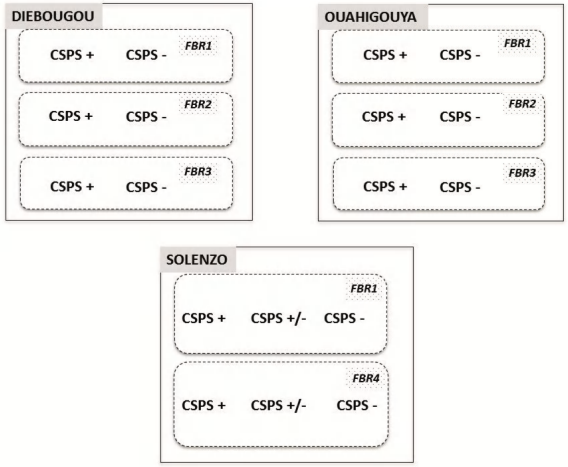


Figure 2. 18 cas de trois districts de santé

Quatre instruments de collecte de données qualitatives ont été utilisés.

1- Des observations (participantes et non participantes) ont été effectuées dans les villages (marchés, réunions communautaires, etc.) pour rédiger une monographie du contexte local et dans les formations sanitaires pour étudier les interactions entre les acteurs et actrices lors des consultations, les visites de contrôle et de supervision, etc. Les résultats des observations ont été consignés dans un journal de bord de chaque assistant.

2- Des entrevues formelles individuelles ont été effectuées à l'aide d'un guide préalablement testé, commun à tous les cas et concernant les dimensions de la théorie de l'intervention. Les personnes ont été sélectionnées après plusieurs jours de présence de terrain en fonction de leur capacité et disponibilité à apporter un éclairage utile sur le FBR et selon le principe d'un échantillonnage qui met l'accent sur la variation et l'intensité au sein d'un même groupe d'acteurs et actrices stratégiques (Palinkas *et al.* 2013) : agent-e-s de santé, agent-e-s de santé communautaires, membres du comité de gestion (COGES), élu-e-s et représentant-e-s associatifs locaux, usagers et populations. Le tableau 1 présente la répartition des 459 entrevues individuelles formelles selon les districts pour ce premier passage (rappelons que trois sont prévus au total afin de réaliser des études de cas longitudinales). En moyenne, nous avons réalisé 22 entrevues par aire de

santé d'une formation sanitaire. La photo 1 illustre une entrevue formelle avec le président d'un COGES.

Entrevue formelle avec le président d'un COGES

3- Des entrevues informelles ont été réalisées : contrairement aux précédentes, il s'agit d'entrevues non planifiées, effectuées sans enregistrement, mais essentielles à la compréhension du FBR dans un contexte d'aide au développement. Les personnes ont été sélectionnées au gré des opportunités (Palinkas *et al.* 2013) et de leur apport pour comprendre la situation étudiée. Des notes ont été prises par la suite afin de conserver les éléments les plus importants de la discussion. La dernière ligne du tableau 1 présente la répartition des 224 entrevues individuelles informelles selon les districts. Les défis liés à cette somme importante d'entrevues sont discutés à la fin du chapitre.

4- De nombreux documents ont été analysés : rapports d'activités et de formation, documents de formations, guide de mise en œuvre, registres de consultations, etc. Des synthèses des éléments importants ont été notées dans le journal de bord des assistant-e-s et chercheuses et chercheurs.

Tableau 1. Nombre d'entrevues par district

	District 1	District 2	District 3
Membres de l'équipe-cadre du district	8	2	5
Professionnel-le de santé	36	46	52
Agent-e-s de santé communautaire	8	12	22
Membres des COGES	14	27	25
Élu-e-s, ONG et leaders locaux	28	20	52
Utilisateurs et utilisatrices des services	17	23	59
TOTAL ENTREVUES FORMELLES	**114**	**130**	**215**
TOTAL ENTREVUES INFORMELLES	**168**	**30**	**26**

Toutes les entrevues ont été intégralement transcrites. Nous avons utilisé le logiciel QDA Miner Lite 4 pour assigner des thèmes aux segments de textes pertinents afin d'organiser et d'extraire les données. La deuxième autrice de ce chapitre a offert une formation sur le logiciel aux membres de l'équipe de recherche, notamment pour tester et adapter les codes ainsi que pour développer une compréhension commune du processus d'analyse. L'analyse a suivi une démarche déductive-inductive en assignant aux données des thèmes prédéfinis sur la base de notre cadre d'analyse (Ritchie et Spenzer 1994), soit la théorie de l'intervention, et en élaborant de nouveaux thèmes en lisant les transcriptions. Les assistant-e-s ont écrit une monographie et un rapport de recherche pour synthétiser les résultats de leurs études de cas (n= 4 assistant-e-s dans les 3 districts). Par la suite, les trois chercheuses et chercheurs responsables des districts ont effectué une analyse transversale intra district (n = 3 rapports). Enfin, le premier auteur a réalisé l'analyse transversale inter districts qui a été validée par les cochercheuses et chercheurs. La figure 3 présente ce processus analytique pour cette première analyse transversale. Le budget global d'une collecte de données (sur les trois prévues) est de l'ordre de 35 000 € comprenant les frais de collecte de données pour les assistants et leurs déplacements, les frais de transcriptions des entrevues, le paiement de leur temps pour les analyses, ainsi que le paiement d'honoraires pour les trois chercheuses et chercheurs séniors du Burkina Faso.

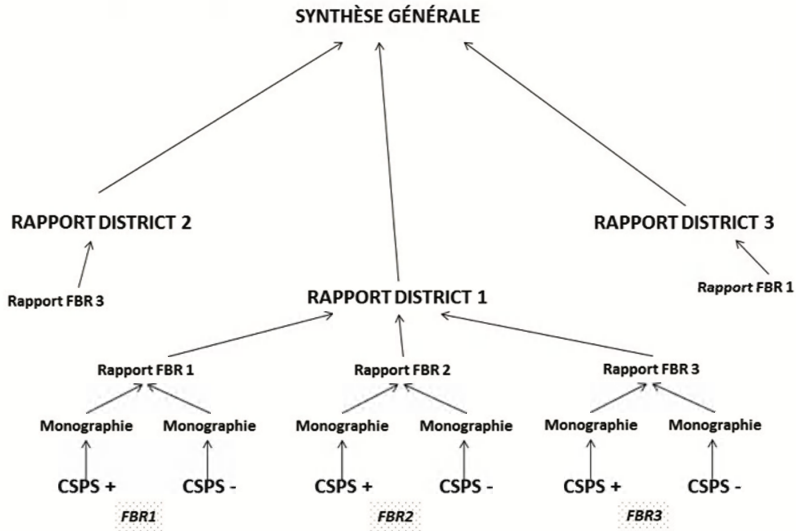


Figure 3. Processus de l'analyse transversale

Quelques résultats

Nous résumons en quelques lignes les principaux résultats de cette recherche qualitative.

Fidèles à une méthode de « formation en cascade » organisée à très grande échelle (y compris par des formations à l'extérieur du pays pour certains cadres), toutes les personnes qui y ont participé sont les mieux informées sur les principes généraux et le mode de fonctionnement du FBR. Toutes les FS ont nommé un point focal FBR (c'est-à-dire une personne référente de l'intervention qui a été formée et qui est censée partager les informations. Mais au-delà de ce cercle, le niveau de connaissance décline pour devenir inexistant pour la population ou les leaders communautaires. Dans certains CSPS, l'octroi des primes au personnel n'a volontairement pas été annoncé aux populations et aux responsables communautaires. La vision globale des personnes rencontrées qui connaissent le FBR reste celle de la présence d'un « *projet* », comme tant d'autres avant et après ce dernier. Ainsi, les répondant-e-s savent qu'il est temporaire, financé et dirigé par l'aide internationale et qu'il peut provoquer de nombreuses logiques d'accaparement, notamment de ressources financières.

La mise en place des agences de contrôle et de vérification et des contrats de performance a pris beaucoup de retard, notamment dans le paiement des subsides et des primes. Ces retards dus à la lourdeur des procédures administratives s'accumulent (recrutement des ACV, signature des contrats, etc.) faisant en sorte que les acteurs et actrices locaux ne savent pas vraiment comment fonctionnent les contrats, les indicateurs ou les plans d'amélioration de la performance (servant à planifier les investissements tirés des subsides du FBR et des autres recettes pour améliorer les services). Les premières vérifications ont partout donné lieu à des discussions d'équipe et à des réajustements rapides des agent-e-s de santé afin que les suivantes soient meilleures pour recevoir le maximum de subsides possibles, car « quand FBR va venir pour l'évaluation, ils sont trop rigoureux, petite erreur, il n'y a pas de pardon dans FBR ! ». Mais la vérification quantitative est souvent source de polémiques (entre vérificateurs et agent-e-s de santé, entre le personnel soignant et le personnel de soutien, et au sein du personnel soignant). En outre, dans quelques CSPS de plusieurs districts, nous avons constaté la préparation d'une mise en scène préalable au contrôle. Ce dernier est toujours annoncé

ou simplement connu, car les informations circulent vite dans ces milieux où le niveau d'interconnaissance est important. De même, certaines données sont inscrites dans les registres, car elles font partie des vérifications, alors que nos observations montrent qu'elles ne sont pas réelles. C'est le cas, par exemple, des prises de constantes des patient-e-s (température, tension, poids) ou du remplissage des partogrammes, que certains agent-e-s de santé nomment des « postogrammes » en privé (c'est-à-dire remplis après coup), tel que cela nous a été rapporté lors des entrevues informelles.

La distribution des primes individuelles est variable entre les centres de santé. Elle dépend du leadership local et des enjeux de pouvoir. Dans certains cas, un manque absolu de transparence est constaté. Ailleurs, mais cela semble être une exception, dans un CSPS performant, le responsable a décidé d'associer tous les agent-e-s de son équipe pour effectuer les calculs des primes en toute transparence.

Dans son design, le FBR n'a pas prévu que des primes soient données aux membres des COGES qui gèrent pourtant le budget, le compte bancaire du CSPS et la mobilisation communautaire. La plupart des formations sanitaires ont respecté cette règle, mais d'autres s'y sont opposées, car elles la trouvaient injuste compte tenu du travail important réalisé par ces membres de la communauté locale. De plus, ce sont eux et elles qui réceptionnent les importantes subventions provenant de l'achat des services par le FBR. Un président de COGES disait à cet égard : « il faut que le FBR songe à motiver aussi les membres du COGES. Comment quelqu'un va aller retirer plus de 400 000 F CFA [600 euros] pour venir donner aux agents qui vont se partager! ».

En somme, les résultats démontrent combien l'adaptation et l'évolution sont la norme et non pas l'exception dans la mise en œuvre des interventions de santé publique, ce qui pose évidemment des défis d'évaluation (Perez *et al.* 2013 ; Ridde et Robert 2014). L'intervention de FBR était très centralisée, son contenu planifié et défini en détail, mais sa rencontre du contexte, des acteurs et actrices et des événements a conduit à ce que tout ne se passe pas comme prévu. Les retards importants des paiements et les enjeux majeurs autour de la distribution des primes sont inquiétants. La faible appropriation locale et le financement externe posent des défis de pérennité, car tant que les responsables gouvernementaux ne se seront pas engagés dans ce type d'intervention et n'auront pas prévu des ressources financières étatiques, la continuité de l'approche est compromise.

Analyse réflexive

La force de notre analyse est qu'elle repose sur une triangulation empirique d'un échantillon très important de données et de cas contrastés, ainsi que sur une collaboration interdisciplinaire assurant la qualité des données. Tous les assistant-e-s ont été formé-e-s et suivi-e-s pour s'assurer de la rigueur de leur travail et la fine connaissance du contexte et des politiques de santé au Burkina Faso par les chercheuses et chercheurs rend leurs interprétations solides. Ce type de processus est relativement rare en Afrique, car on demande souvent aux assistant-e-s de « simplement » collecter les données qui sont ensuite analysées par les chercheuses et chercheurs du pays, voire pire uniquement par les chercheuses et chercheurs du Nord (Ouattara et Ridde 2013). Cela n'a évidemment pas été facile et le contexte du Burkina s'y prête peut être mieux qu'ailleurs, car nous y travaillons depuis longtemps, disposant ainsi d'un réseau de chercheuses et chercheurs de longue date, en général rattaché-e-s à des institutions. Ils ont ainsi un large réseau d'étudiant-e-s au sein duquel nous avons été en mesure de trouver des personnes compétentes. Au Mali où nous tentons d'organiser le même processus, la situation est plus complexe, les ressources humaines semblant plus rares et nos collègues chercheuses et chercheurs moins associés à des programmes de formation pour disposer d'un réseau d'étudiant-e-s.

Les résultats ont été largement partagés et validés auprès de nombreuses parties prenantes du FBR dans le pays. Ainsi, nous pensons qu'ils peuvent être transférables à l'ensemble des districts du pays impliqués dans la mise en œuvre du FBR.

De plus, les résultats seront certainement utiles pour mieux comprendre l'évaluation d'impact. Depuis le début de cette recherche sur la mise en œuvre, nous entretenons des liens étroits avec les chercheuses et chercheurs de l'équipe de l'évaluation d'impact. Notre expertise sur la mise en œuvre s'est avérée utile pour aider cette équipe à mieux comprendre le déroulement réel de l'intervention (au-delà des manuels d'intervention initiaux) et à adapter une partie de leurs questionnaires standardisés de fin afin d'améliorer leur pertinence dans le contexte local. Cette collaboration représente une valeur ajoutée pour nos recherches respectives et le développement de connaissances en faisant le pont entre les approches qualitatives et quantitatives.

Dans le cas de cette étude de mise en œuvre, le défi majeur reste la quantité très (trop ?) importante de données qualitatives et donc son corollaire qui a été la durée de l'analyse. Il a fallu plus d'un an pour disposer des premiers rapports d'analyse des cas permettant de démarrer une analyse collective et transversale des 459 entrevues formelles et 224 entrevues informelles provenant des 21 formations sanitaires. L'utilisation d'un processus pyramidal par lequel chaque membre de l'équipe a analysé et synthétisé des données à un niveau différent (CSPS, district et général) a facilité la gestion des données en répartissant la charge de travail. En ayant recours à une telle quantité de données collectées à travers différents sites, nous espérons que les résultats pourront expliquer l'hétérogénéité de la mise en œuvre, en plus de convaincre un large public cible de la validité des résultats, incluant des décideuses et décideurs et certains chercheuses et chercheurs pour qui la taille de l'échantillon demeure essentielle. En effet, notre expérience dans la région montre que, bien souvent, les décideuses et décideurs discréditent les recherches qualitatives à cause de la faible taille des échantillons, ne maitrisant pas les notions de saturation empirique ou n'ayant pas de connaissances des critères de scientificité en qualitatif (Laperrière 1997). Dans ce contexte, comment trouver un bon équilibre entre le besoin de données suffisantes pour répondre à une question de recherche complexe autour d'une intervention de grande ampleur et disposant de multiples composantes (bras), les défis éthiques de rencontrer autant de personnes et de mobiliser une équipe aussi importante, et les demandes des décideuses et décideurs d'une plus grande validité externe des évaluations de programmes ? La présente analyse de la mise en œuvre est une tentative de réconciliation de ces mondes sans pour autant avoir la prétention de clore la discussion.

Un autre défi dans le contexte de cette collaboration internationale a été de maintenir une communication dynamique entre les différents membres de l'équipe, dispersés dans différentes régions au Burkina Faso et à travers le monde. Nous avons eu écho de malentendus, notamment en lien avec les rôles de chacun des membres de l'équipe, les méthodes de collecte de données, les délais de l'étude et les modalités de rémunération des assistant-e-s de recherche. La distance, la taille de l'équipe et les problèmes de réseaux de communication, ainsi que les coûts associés, ont certainement contribué à l'ampleur du défi. En outre, dans le contexte d'une collaboration internationale, les retards dans les analyses et les partages des rapports

ne sont jamais simples à gérer, car il faut à la fois être un chercheur ou chercheuse (du Nord qui plus est !) et travailler avec des collègues, mais en même temps régler les questions financières et de délais pour disposer de produits suffisamment solides. Ces défis sont bien connus (Ouattara et Ridde 2013).

Une des forces de notre démarche a été l'organisation d'une formation sur l'utilisation d'un logiciel d'analyse qualitative gratuit (QDA miner lite) pour les membres de l'équipe (voir Photo 2). Peu commune au Burkina Faso, cette formation a permis de renforcer les capacités des assistant-e-s pour cette recherche ainsi que pour leur future carrière dans le domaine. Il s'agissait aussi d'une occasion de les intégrer dans l'ensemble du processus et de ne pas en faire de simple collecteurs de données, comme c'est bien souvent le cas, ce qui pose des problèmes importants pour les recherches qualitatives. Plusieurs participant-e-s ont exprimé leurs intérêts et gratitude à cet égard.

Une difficulté principale lors de la formation a été de gérer les différents niveaux de compétence en informatique des participant-e-s. Nous avons dû trouver un rythme convenant à l'ensemble du groupe. Une autre des difficultés rencontrées lors de la formation a été de se mettre d'accord sur les définitions et la pertinence de certains codes d'analyse. Par exemple, nous avons eu un débat sur la création de codes distinct pour « utilité du FBR » et « pertinence du FBR ». Ces discussions sont importantes pour développer une compréhension commune du processus d'analyse, mais peuvent être laborieuses lorsqu'il y a une centaine de codes à élaborer. Par ailleurs, l'utilisation d'un logiciel d'analyse qualitative pour organiser les données a facilité la réalisation d'analyses secondaires pour étudier des thématiques spécifiques (par exemple, la qualité des soins, la motivation des acteurs et actrices) subséquemment.

Formation sur l'utilisation d'un logiciel d'analyse qualitative gratuit

Malgré les efforts de certains membres de l'équipe, la centralisation des nombreux documents de la recherche (enregistrements des entrevues, notes d'observations, transcriptions des entrevues, données codifiées) a constitué un défi majeur pour cette étude. La centralisation des documents est importante afin d'éviter la perte de données, de vérifier des

données brutes et de permettre des analyses secondaires. Toutefois, la centralisation s'est avérée difficile tout au long de la recherche, notamment en raison de la dispersion géographique des assistant-e-s de recherche et de la mauvaise connexion internet qui nuit au partage rapide de fichiers lourds. Plus de trois ans après le début de cette recherche, nous continuons de solliciter la centralisation de tous les documents.

Cette étude s'est inscrite dans le courant de l'évaluation fondée sur la théorie de l'intervention. Ainsi, c'est plus la théorie de l'intervention qui a guidé la collecte et l'analyse des données que les cadres ou théories de la science de la mise en œuvre. Cela étant dit, les concepts de ce champ ont largement été utilisés lors de nos réflexions et analyses, par exemple la notion de contexte, le rôle des acteurs et actrices et des enjeux de pouvoir, les déficits de mise en œuvre, le rôle central des acteurs et actrices de première ligne compris comme les véritables « faiseurs » de politique (Lipsky 2010; Pfadenhauer *et al.* 2015; Pressman et Wildavsky 1984 ; Shoveller *et al.* 2015), etc. Au Mali cependant, nous nous sommes largement inspirés d'un cadre conceptuel spécifique pour l'analyse de la mise en œuvre afin d'étudier l'intervention de FBR dans une région du pays (Damschroder *et al.* 2009). Nous serons en mesure d'en tirer prochainement des leçons, notamment dans les défis de son usage par des anthropologues qui sont peu habitués, et c'est normal, à trop encadrer les dimensions qu'ils et elles étudient. Mais face à la complexité des interventions, au besoin d'études de cas multiples pour augmenter la validité externe des résultats demandée par les décideuses et décideurs plus habitués aux données quantitatives et à la quantité phénoménale de données qualitatives nécessaires, le passage par ce type de cadre conceptuel et d'analyse assistée par ordinateur est peut-être incontournable. Il devient essentiel de tenter d'innover dans ce domaine (Belaid *et al.* 2016).

Le chapitre présente des résultats d'une collecte de données transversale, traduisant la situation à un moment donné et non son évolution (bien que, même au cours des 12 mois de l'étude, des adaptations aient été largement constatées). Mais nous souhaitions partager les résultats sans attendre les autres phases tant les recherches sur le sujet sont rares en Afrique de l'Ouest francophone. Les prochaines phases de notre programme de recherche permettront de mieux comprendre les changements dans l'intervention, son adaptation, et la manière dont le contexte et les acteurs et actrices agissent dans le temps. Les études de cas longitudinales sont

en effet essentielles à la compréhension des politiques, mais difficiles à organiser (Saetren 2014; Ridde *et al.* 2014; Hill et Hupe 2014).

L'évaluation des processus de mise en œuvre des interventions est essentielle, non seulement car elle permet de comprendre les effets (ou leur absence) des actions entreprises, mais surtout, car elle rend compte de la complexité des processus, du rôle des acteurs et actrices et du contexte. Ce chapitre présente un exemple de la manière dont la théorie de l'intervention, et plus largement des cadres conceptuels de la science de la mise en œuvre, peuvent être mobilisés à cet égard. Assurément, nous avons collectivement besoin de plus d'évaluations de la mise en œuvre des interventions en santé mondiale, mais aussi de meilleures évaluations, notamment avec l'utilisation de cadres conceptuels et de données qualitatives.

Références clés

Patton, M. Q. (2008). Implementation Evaluation: What happened in the program? *Utilization-focused Evaluation*. 4th Edition. Saint-Paul: SAGE Publications.

Patton livre un plaidoyer convaincant sur l'importance d'étudier la mise en œuvre d'interventions. L'auteur raconte l'histoire farfelue d'un groupe d'évaluateurs et évaluatrices qui a évalué les effets d'une intervention qui n'avait jamais réellement été implantée, tirant de mauvaises conclusions sur son inefficacité. L'auteur présente cinq types d'évaluation : l'évaluation des efforts, le monitorage, l'évaluation des processus, l'évaluation des composantes et la spécification du traitement. Le chapitre termine avec une liste d'exemples de questions d'évaluation en lien avec la mise en œuvre.

Damschroder, L. *et al.*, (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science*, 4(50).

Cet article présente un modèle conceptuel, le Consolidated Framework For Implementation Research, issu de différentes théories. Le modèle est composé de cinq domaines : les caractéristiques de l'intervention, le milieu externe, le milieu interne, les caractéristiques des individus et le processus d'implantation. Les auteurs présentent des définitions utiles pour les différents concepts et leurs construits. Plusieurs autres ressources sont disponibles sur le site internet suivant : https://cfirguide.org/

Olivier de Sardan, J.-P. (2008). *La rigueur du qualitatif. Les contraintes empiriques de l'interprétation socioanthropologique*. Louvain-La-Neuve: Bruylant.

C'est un livre à lire absolument! Il permet de saisir l'importance de s'insérer de façon prolongée dans un milieu pour comprendre la mise en œuvre d'une intervention. Il permet aussi de s'outiller pour faire face aux enjeux de l'observation participative. Le livre aborde plusieurs questions importantes: Sur quelles procédures repose l'observation? Quelles sont les conditions de la rigueur dans les approches qualitatives? Quelle place accorder à la subjectivité du chercheur ou de la chercheuse? Peut-on minimiser les pièges de la surinterprétation ou les biais des idéologies scientifiques?

Ridde, V., Yaogo, M., Somé, P.-A., Turcotte-Tremblay, A.-M., (2017). Twelve months of implementation of health care performance-based financing in Burkina Faso: A qualitative multiple case study. *The International Journal of Health Planning and Management*. 33 (1), e153-e167.

Cet article présente notre étude sur la mise en œuvre du financement basé sur les résultats au Burkina Faso. Il permettra de voir un exemple concret du type d'article scientifique que l'on peut publier en étudiant la mise en œuvre d'une intervention complexe en santé mondiale. L'article contient une référence à notre protocole complet qui pourra vous servir d'exemple pour développer un protocole visant à comprendre comment une intervention fonctionne, pour qui et dans quels contextes réels.

Références

Belaid, L., Bodson, O., Ridde, V., Samb, O. et Turcotte-Tremblay, A.-M. (2016). Innovations qualitatives dans l'évaluation des interventions en santé. Dans J. Kivits, F. Balard, C. Fournier et M. Winance (dir.), *Les recherches qualitatives en santé* (p. 260-276). Malakoff : Armand Colin.

Béland, D. et Ridde, V. (2016). Ideas and policy implementation: Understanding the resistance against free health care in Africa. *Global Health Governance*, 10(3), 9-23.

Breton, E. et De Leeuw, E. (2011). Theories of the policy process in health promotion research: a review. *Health Promotion International*, 26(1), 82-90.

Cox, R. H. (2001). The social construction of an imperative: Why welfare

reform happened in Denmark and the Netherlands but not in Germany. *World Politics*, 53(3), 463-498.

Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A. et Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation ccience. *Implementation Science*, 4, 50-64. https://doi.org/10.1186/1748-5908-4-50

Dobson, D. et Cook, T. D. (1980). Avoiding type III error in program evaluation: Results from a field experiment. *Evaluation and Program Planning*, 3(4), 269-276.

Durlak, J. A. et DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41(34), 327-350. https://doi.org/10.1007/s10464-008-9165-0

Durlak, J. A. (2015a). What everyone should know about implementation. Dans J. A. Durlak, C. E. Domitrovich, R. Weissberg et T. Gullotta (dir.), *Handbook of social and emotional learning: research and practice* (p. 395-405). New York: The Guilford Press.

Durlak, J. A. (2015b). Studying program implementation is not easy but it is essential. *Prevention Science*, 16(8), 1123-1127. https://doi.org/10.1007/s11121-015-0606-3

Franks, B. et Schroeder, J. (2013). Implementation science : What do we know and where do we go from here?. Dans A. Metz et I. Martinez-Beck (dir.), *Applying implementation science in early childhood programs and systems* (p. 5–21). Baltimore: Paul H. Brookes Publishing Company.

Gilson, L. et Raphaely, N. (2008). The terrain of health policy analysis in low and middle income countries: a review of published literature 1994–2007. *Health Policy and Planning*, 23(5), 294-307.

Hill, M. J. et Hupe, P. L. (2014). *Implementing public policy: an introduction to the study of operational governance* (3e éd.). Thousand Oaks, CA: Sage Publications.

Laperrière, A. (1997). Les critères de scientificité des méthodes qualitatives. Dans J. Poupart, J-P. Deslauriers, L-H. Groulx, A. Laperrière, R. Mayer

et A.P. Pires (dir.), *La recherche qualitative : enjeux épistémologiques et méthodologiques* (p. 365-389). Montréal : Gaëtan Morin.

Lipsky, M. (2010). *Street-level bureaucracy: Dilemmas of the individual in public services*. New York : Russel Sage Foundation.

Milstein, R. et Schreyoegg, J. (2016). Pay for performance in the inpatient sector: A review of 34 P4P programs in 14 OECD countries. *Health Policy*, 120(10), 1125-1140.
https://doi.org/10.1016/j.healthpol.2016.08.009

Nilsen, P., Ståhl, C., Roback, K. et Cairney, P. (2013). Never the twain shall meet? – a comparison of implementation science and policy implementation research. *Implementation Science*, 8, 63.
https://doi.org/10.1186/1748-5908-8-63

Olivier de Sardan, J.-P et Ridde, V. (2015). Public policies and health systems in Sahelian Africa: Theoretical context and empirical specificity. *BMC Health Services Research*, 15(Suppl. 3), S3.
https://doi.org/10.1186/1472-6963-15-S3-S3

Ouattara, F. et Ridde, V. (2013). Expériences connues, vécues… mais rarement écrites : à propos des relations de partenariat Nord-Sud. *Nouvelles Pratiques Sociales*, 25(2), 231-246.

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N. et Hoagwood, K. (2013). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(5), 533-544.
https://doi.org/10.1007/s10488-013-0528-y

Patton, M. Q. (2010). *Developmental evaluation : Applying complexity concepts to enhance innovation and use*. Lieu d'édition: The Guilford Press.

Paul, E., Sossouhounto, N. et Eclou, D. S. (2014). Local stakeholders' perceptions about the introduction of performance-based financing in Benin: A case study in two health districts. *International Journal of Health Policy and Management*, 3(4), 207-214.
https://doi.org/10.15171/ijhpm.2014.93

Perez, D., Lefevre, P., Castro, M., Toledo, M. E., Zamora, G., Bonet, M. et Van der Stuyft, P. (2013). Diffusion of community empowerment strategies for Aedes aegypti control in Cuba: A muddling throughexperience. *Social*

Science and Medicine, 84, 44-52.
https://doi.org/10.1016/j.socscimed.2013.02.003

Peters, D. H., Adam, T., Alonge, O., Agyepong, I. A. et Tran, N. (2013). Implementation research: what it is and how to do it. *British Medical Journal*, 347, f67-53.
https://doi.org/10.1136/bmj.f6753

Peters, D. H., El-Saharty, S., Siadat, B., Janovsky, K. et Vujicic, M. (dir.). (2009). *Improving health service delivery in developing countries: from evidence to action*. Washington: World Bank.

Petticrew, M., Cummins, S., Ferrell, C., Findlay, A., Higgins, C., Hoy, C., … et Sparks, L. (2005). Natural experiments: an underused tool for public health?. *Public Health*, 119(9), 751-757.

Pfadenhauer, L. M., Mozygemba, K., Gerhardus, A., Hofmann, B., Booth, A., Bakke, K. B., … et Rehfuess, E. A. (2015). Context and implementation: A concept analysis towards conceptual maturity. *Zeitschrift Für Evidenz, Fortbildung Und Qualität Im Gesundheitswesen*, 109(2), 103-114.
https://doi.org/10.1016/j.zefq.2015.01.004

Pressman, J. L. et Wildavsky, A. (1984). *Implementation: How great expectations in Washington are dashed in Oakland* (3e éd.). Berkeley, Los Angeles, London : University of California Press.

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., … et Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health*, 38(2), 65-76.
https://doi.org/10.1007/s10488-010-0319-7

Ridde, V. (2009). Policy implementation in an African state : an extension of the Kingdon's multiple-streams approach. *Public Administration*, 87(4), 938-954.

Ridde, V., Delormier, T. et Gaudreau, L. (2007). Evaluation of empowerment and effectiveness: Universal concepts ?. Dans D. V. McQueen et C. M. Jones (dir.), *Global Perspectives on Health Promotion Effectiveness* (vol. 1, p. 389-403). New York: Springer Science et Business Media.

Ridde, V. et Robert, É. (2014). *Real World Evaluation Strategies*. Oxford University Press.

http://www.oxfordbibliographies.com/display/id/obo-9780199756797-0140

Ridde, V., Turcotte-Tremblay, A.-M., Souares, A., Lohmann, J., Zombré, D., Koulidiati, J., ... et De Allegri, M. (2014). Protocol for the process evaluation of interventions combining performance-based financing with health equity in Burkina Faso. *Implementation Science*, 9(1), 149-160. https://doi.org/10.1186/s13012-014-0149-1

Ritchie, J. et Spenzer, L. (1994). Qualitative data analysis for applied policy research. Dans A. Bryman et R. G. Burgess (dir.), *Analyzing qualitative data* (p. 173-194). Londres and New York: Routledge.

Saetren, H. (2014). Implementing the third generation research paradigm in policy implementation research: An empirical assessment. *Public Policy and Administration*, 29(2), 84-105. https://doi.org/10.1177/0952076713513487

Shoveller, J., Viehbeck, S., Di Ruggiero, E., Greyson, D., Thomson, K. et Knight, R. (2015). A critical examination of representations of context within research on population health interventions. *Critical Public Health*, *volume*(numéro), 114. https://doi.org/10.1080/09581596.2015.1117577

Siron, S., Dagenais, C. et Ridde, V. (2015). What research tells us about knowledge transfer strategies to improve public health in low-income countries: A scoping review. *International Journal of Public Health*, 60(7), 115. https://doi.org/10.1007/s00038-015-0716-5

Turcotte-Tremblay, A.-M., Spagnolo, J., De Allegri, M. et Ridde, V. (2016). Does performance-based financing increase value for money in low- and middle- income countries?: A systematic review. *Health Economics Review*, 6(1). https://doi.org/10.1186/s13561-016-0103-9

United Nations Secretary-General. (2010). *Global Strategy for Women's and Children's health. Geneva* : World Health Organization. http://www.who.int/pmnch/knowledge/publications/fulldocument_globalstrategy/en/.

Van Belle, S., van de Pas, R. et Marchal, B. (2017). Towards an agenda for implementation science in global health: there is nothing more practical

than good (social science) theories. *BMJ Global Health*, 2(2), e000181. https://doi.org/10.1136/bmjgh-2016-000181

Walt, G., Shiffman, J., Schneider, H., Murray, S. F., Brugha, R. et Gilson, L. (2008). "Doing" health policy analysis: Methodological and conceptual reflections and challenges. *Health Policy and Planning*, 23(5), 308-317.

Weiss, C. H. (1998). *Evaluation : methods for studying programs and policies* (2e éd.). Upper Saddle River, NJ : Prentice Hall.

Witter, S., Fretheim, A., Kessy, FL. et Lindahl, AK. (2012). Paying for performance to improve the delivery of health interventions in low- and middle-income countries. *Cochrane Database of Systematic Review*, 2, CD007899.

Yin, R.K., et Ridde, V. (2012). Théorie et pratiques des études de cas en évaluation. Dans V. Ridde et C. Dagenais (dir.), *Approches et pratiques en évaluation de programme* (2e éd., p. 177-193). Montréal : Presses de l'Université de Montréal.

Zombré, D., Turcotte-Tremblay, A.-M., Sangaré, A., Zongo, S., Somé, P.-A., Coulibaly, A., ... et Ridde, V. (2016). *Méthode mixte et participative de sélection d'études de cas multiples pour l'évaluation d'intervention de santé au Burkina Faso et au Mali : De l'éthique à l'émique.* Communication présentée au Congrès de l'Association francophone pour le savoir, Montréal.
https://docs.google.com/viewerng/viewer?url=http://www.equitesante.org/wp-content/uploads/2016/05/Poster.David_.ACFAS_.pdfethl=fr.

Résumé / Abstract / Résumé en moore / Résumé en dioula

Plus de recherches sur la mise en œuvre sont nécessaires pour comprendre comment les interventions fonctionnent dans leur contexte réel et comment elles évoluent dans le temps. L'objectif de ce chapitre est de présenter l'analyse de la mise en œuvre d'une intervention complexe, intitulée le financement basé sur les résultats, au Burkina Faso. Nous avons effectué une étude de cas multiples (18 formations sanitaires) et contrastés avec plusieurs niveaux d'analyse imbriqués. Les données empiriques proviennent d'observations, d'entrevues informelles et formelles et de

documents. Pour faciliter l'analyse d'une quantité importante de données, nous avons utilisé un processus pyramidal par lequel chaque membre de l'équipe a analysé et synthétisé des données à un niveau différent (centre de santé, district et général). Ainsi, les assistants de recherche ont participé à l'ensemble du processus d'analyse. L'organisation d'un atelier sur l'utilisation d'un logiciel d'analyse qualitative accessible gratuitement a permis de développer une compréhension commune du processus d'analyse. Les résultats de la recherche montrent l'importance d'ouvrir la « boite noire » pour comprendre la mise en œuvre des interventions de santé publique en Afrique et mettre au jour leurs adaptations selon les contextes et les acteurs et actrices sociaux.

***

More research on implementation is needed to understand how interventions work in their real context and how they evolve over time. The objective of this chapter is to present the process of analyzing the implementation of a complex intervention, called performance-based financing, in Burkina Faso. We conducted a multiple and contrasted case study (18 healthcare facilities) with several levels of analysis. The empirical evidence comes from observations, informal and formal interviews and documents. To facilitate the analysis of a large amount of data, we used a pyramid process whereby each team member analyzed and synthesized data at a different level (healthcare facility, district, general). Thus, the research assistants participated in the whole process of analysis. The organization of a workshop on the utilization of a free qualitative analysis software allowed the development of a common understanding of the analysis process. The results of the study highlight the importance of opening the "black box" to understand the implementation of interventions in global health and to shed light on their adaptations according to the contexts and the social actors.

***

Sen na yıle ti d wʊm sõngrã na-kendr ã võor la a sen teesde, tʊʊmdã siglgl toga ne vaɛɛsg tı seke. Zãms- kãnga yegre yaa sen na yıl n wilg bumb ning sen be sõng-kaseng sen sigend tʊʊm- biis zug Burkina faaso ka wã . Tõnd talla bʊgsg sen keed ne yɛl buud toor-toor logtoy-yi ziis piig la nii(18) n tı yek ne vaɛɛsg bɛd a taaba.
Vaɛɛs kãng biisã tuuda ne ziig pʊge gesgo, sõsg ne neb la sɛb wʊsg me karengo. Sen na yıl tı veenem zind vaɛɛs kaseng kãnga zugu faag-faaga tõnd

kɩtam ti tʊm- tʊmdb sen be yingr n tãag tengr kõ b tagsgo (logto-yi la bɛd la boonese). Rend bãngdba taoor dãmba zĩnda tʊʊmda sıngre n tı tãag a baasgo. Tigsg sen n kõ zãmsg tʊʊma-teed zug zaalem sõngam tı bãng sõng paam vaɛɛsgã zugu. Tʊʊmda biis wilgam tı yaa tılae ti b pak koglg buud sen na n sõng ninsables zamaana laafı wa koglgo, la b le wilg sen tokd ne tʊm-tʊmdba fãa raabo.

***

Fɛɛrɛ yaala ka taga lɔɲni fanfɛ min be na ni kow famuyali basigili ye, ka kan ka sira damanin ta min sababura famuyali kɛrɛn kɛrɛn be dafali sɔ̰rɔ̰, Nin baara nin kunsiyɔ̰rɔ̰ ye, ka se ka sɛgɛsɛgɛli kɛ ka taga kow gwɛlɛma fɛrɛ fanfɛ a be fɔ̰ min ma ko harajigɛ jaga bɔ̰ n'o basigira tɔɲɔ̰ sɔ̰rɔ̰li kan Burukin Faso yan.

An ye sɛgɛsɛgɛli kɛ koo caaman le kan dɔgɔ̰tɔ̰rɔ̰so tan ani segi(18) le kan minw bɔ̰len tɛ yɔɲgɔɲ ma.ɛɛgɛsɛgɛli baara kɛra ka taga baro ani sɛbɛn damani kɔɲɔ̰ ka se ka lɔɔ̰ni sɔ̰rɔ̰ o senkɔ̰rɔ̰,Ka se ka famuyali sɔ̰rɔ̰ o sɛgɛsɛgɛli kan nin kan, an ye tagamasiyɛn jɔɲjɔɲ ta minw sababababura baarakɛla minw bɛɛ tun be jɛnkulu nin kɔɲɔ̰ ka se k'o lɔyɔ̰rɔ̰ baara kɛ.(Dɔgɔ̰tɔ̰rɔ̰sodennin, a n'afabaw). ka tugu minw tun be lɔɔ̰nikɛlagana nunu dɛmɛ ka baara sɛgɛsɛgɛli nin kɛ, o fana ye tali kɛ baara mumɛ bɛɛ ra.

Yɔɲgɔɲyebaara dɔ̰ sigi ra ka se ka lɔɲni sɛgɛsɛgɛli nin baarakɛminanw labila gwansan min senkɔ̰rɔ̰ bɛɛ ka se ka famuyali kelen sɔ̰rɔ̰ a ra ani o hakilinata ka kɛ kelen ye. Baara nin ka fɛrɛ kɛra sababu ye ka gondon damani famuyali ngɔgɔya ka taga jamakulu ka hɛrɛ kow lasigili kan farafina jamana kɔɲɔ̰,Baarakɛminaw min bilara lɔɲnikɛla bolo dira o ma gwansan. A ti san ni wari ye.ɛɛgɛsɛgɛli nin tɔɲɔ̰ lɔɲnikɛlaw ka min basigi ka taga hɛrɛko fanfɛ farafina bolo kɔɲɔ̰ ka se bɛn jamakulu ka hakilinata ye ,ani k'o fɛrɛw kɛrɛnkɛrɛn ta ka k'o ka sininyasigi jagabɔ̰ wala lɔɲnikɛlaw ya yira cogomina.

Nous remercions Alain Ouédraogo, Ahmed Sie Barro, Mme Namoano Solange et Abbe Anselme Yanogo pour les traductions en langue moore et dioula.

***

Valéry Ridde est directeur de recherche au CEPED[4], une Unité Mixte de Recherche associant l'Université Paris Descartes et l'Institut de Recherche pour le Développement (IRD). Il est chercheur régulier de l'Institut de

4. http://www.ceped.org

recherche en santé publique de l'Université de Montréal (IRSPUM) et chercheur associé au Centre interdisciplinaire de recherche en développement international et société (CIRDIS) de Montréal. Il est rédacteur adjoint des revues *BMJ Global Health* et *Global Health Promotion*, membre du comité de rédaction de la *Revue Canadienne d'Évaluation de Programme* et de *Qualitatif! Revue africaine de sciences sociales*. Ses travaux de recherche portent sur la couverture universelle en santé, le financement des services de santé, l'évaluation de programme, les politiques publiques de santé et la promotion de la santé.

Anne-Marie Turcotte-Tremblay est candidate au doctorat en santé publique avec une spécialisation en organisation des soins à l'École de santé publique de l'Université de Montréal (ESPUM). Elle est également affiliée à l'Institut de recherche en santé publique de l'Université de Montréal (IRSPUM). Sa thèse de doctorat porte sur les conséquences non intentionnelles du financement basé sur les résultats combinés à des mesures de protection sociale au Burkina Faso. Anne-Marie est boursière des Instituts de recherche en santé du Canada (IRSC). Elle a coordonné plusieurs recherches internationales en santé des populations. Anne-Marie s'intéresse à l'évaluation d'interventions visant à réduire les inégalités sociales de santé.

Citation

Valéry Ridde et Anne-Marie Turcotte-Tremblay. (2019). L'analyse des processus de mise en œuvre. Une intervention complexe au Burkina Faso : le financement basé sur les résultats. In *Évaluation des interventions de santé mondiale. Méthodes avancées*. Sous la direction de Valéry Ridde et Christian Dagenais, pp. 355-381. Québec : Éditions science et bien commun et Marseille : IRD Éditions.

15. L'évaluation de la fidélité d'implantation

Un projet de distribution d'omble chevalier aux femmes enceintes du Nunavik

LARA GAUTIER, CATHERINE M. PIRKLE, CHRISTOPHER FURGAL ET MICHEL LUCAS

L'évaluation de la fidélité d'implantation

Définition de la méthode

Le cadre évaluatif de la fidélité d'implantation développé par Carroll et ses collègues (2007) met en lumière l'écart entre les intentions initiales des concepteurs et conceptrices de projets d'intervention et les moyens par lesquels ceux-ci sont mis en œuvre. Plus cet écart est large, moins la « fidélité d'implantation » est élevée. Les caractéristiques du projet et les éléments « modérateurs » (par exemple les stratégies de facilitation, la qualité de la prestation, la réceptivité des participant-e-s à l'intervention) peuvent influer sur le degré de fidélité d'implantation.

Force de la méthode

- Aide les chercheuses et chercheurs et les décideuses et décideurs à améliorer l'implantation de projets ou programmes d'intervention.
- Permet d'évaluer de façon systématique les obstacles et les facilitateurs à l'implantation du projet. Ceci est particulièrement pertinent dans le domaine de la santé des populations vulnérables en raison des niveaux de complexité élevés du contexte d'implantation.

Défis de la méthode

- Attention insuffisamment portée au contexte sociopolitique et à l'environnement dans lequel l'intervention est implantée (Hasson 2010). Cette limite a émergé lors des entretiens, à travers la manifestation de préoccupations au sujet de la viabilité financière et environnementale du projet.
- Une approche « individualiste » qui ignore le contexte humain et les liens entre les personnes impliquées. Ceci est d'autant plus problématique dans l'évaluation d'interventions mises en œuvre dans le contexte autochtone, où la frontière entre l'individu et la communauté est ténue.

Un village du Nunavik isolé en période d'hiver, 2014. Crédits : Isabelle Veillette.

Avec le développement du champ de l'évaluation des processus, de nouvelles approches émergent pour explorer la boite noire de l'implantation. Ces approches aident à comprendre les manières dont les interventions atteignent des résultats efficaces. Dans le domaine de la santé des Autochtones au Canada, il y a un intérêt croissant pour mieux comprendre dans quelle mesure les interventions sont mises en œuvre et comment les améliorer. De nouvelles méthodes ont notamment émergé pour explorer en profondeur le degré d'implantation de programmes ou « projets » d'intervention. La présente étude qualitative utilise un cadre novateur pour l'évaluation d'implantation : celui de Carroll et ses collègues (2007). Les auteurs et autrice suggèrent que les moyens par lesquels les programmes sont mis en œuvre peuvent différer considérablement des intentions initiales des concepteurs et conceptrices. Cet écart représente le degré de « fidélité d'implantation » qui indique que plus l'écart entre les intentions et les moyens de mise en œuvre est important, moins la fidélité d'implantation est élevée.

Conceptuellement, selon Carroll et ses collègues (2007), la fidélité d'implantation inclut les caractéristiques d'un projet (contenu, couverture, fréquence et durée), ainsi que les modérateurs qui peuvent influer sur le degré d'implantation (complexité du projet, stratégies de facilitation, qualité de la prestation et réceptivité des participant-e-s au projet). Nous avons adapté ce cadre théorique pour évaluer la fidélité d'implantation dans le contexte du projet de distribution d'omble chevalier au Nunavik. Il est composé de six éléments-thèmes clés. Les cinq premiers se réfèrent aux différents modérateurs, alors que le sixième est celui qui correspond aux indicateurs d'un projet :

- l'impact attendu du projet
- la complexité inhérente à l'intervention (notamment au plan logistique)
- les stratégies de communication et coordination pour implanter le projet

- la qualité de la prestation de l'intervention (informations relatives à la différence entre ce que les bénéficiaires sont censés recevoir et ce qu'ils et elles obtiennent réellement)
- la réceptivité au projet, c'est-à-dire la manière dont les acteurs et actrices impliqués dans le projet apprécient celui-ci. La réceptivité est analysée auprès des bénéficiaires, mais aussi auprès des personnes ou groupes de personnes chargées de l'implantation du projet, auprès des organisations impliquées : dans quelle mesure leurs caractéristiques structurelles ont-elles une influence (leadership, ressources humaines, reddition des comptes, financement, etc.)?
- les liens de correspondance entre les caractéristiques du projet (couverture du projet, dosage, fréquence, durée, etc.) et la réalité.

Ces six thèmes constituent les éléments de base de mesure de la fidélité d'implantation du projet évalué dans ce chapitre.

Contexte et projet évalué

Description du contexte où se déroule le projet

Des bouleversements socioéconomiques et environnementaux ont affecté les communautés autochtones du Nord canadien au cours des 60 dernières années. En particulier, ces bouleversements ont eu un impact négatif sur de nombreux systèmes alimentaires basés sur les produits locaux issus de la chasse, de la pêche et de la cueillette. Aujourd'hui, parmi les populations autochtones des pays industrialisés, les communautés inuites du Canada ont les plus hauts niveaux d'insécurité alimentaire (Council of Canadian Academies 2014). En outre, l'exposition aux contaminants environnementaux, tels que le mercure, menace la sécurité alimentaire de ces populations (Pirkle, Muckle et Lemire 2016; Van Oostdam et al. 2005). Compte tenu d'un régime alimentaire incluant la viande de mammifères marins, elles sont exposées de manière disproportionnée au mercure par rapport aux Canadiens du Sud (Pirkle, Muckle et Lemire 2016). Dangereux pour le fœtus en développement, le mercure a aussi des répercussions négatives à long terme sur le développement de l'enfant (Boucher *et al.* 2012; Jacobson *et al.* 2015; Pirkle, Muckle et Lemire 2016).

Les activités de promotion de la santé fondée sur les systèmes alimentaires des Inuits et la réduction de l'exposition aux contaminants environnementaux sont nécessaires. Certains aliments locaux (par exemple la viande de béluga) contiennent des niveaux élevés de nutriments, mais peuvent aussi présenter des concentrations élevées en mercure (Lemire *et al.* 2015). Une possibilité d'intervention est de promouvoir la consommation d'aliments locaux inuits qui sont riches en nutriments mais pauvres en contaminants environnementaux. L'omble chevalier (*Salvelinus alpinus*), un poisson d'Amérique du Nord, constitue une source d'alimentation correspondant précisément à ces critères.

Prospectus distribué dans le cadre de la journée des aliments traditionnels, 2014. Crédits photo : Isabelle Veillette.

Description de la logique du projet

Au Nunavik, une région du Nord-du-Québec (Canada), il existe un projet communautaire qui vise à améliorer la qualité nutritionnelle de l'alimentation, à réduire l'insécurité alimentaire et à minimiser l'exposition aux contaminants environnementaux. Dans sa conception, ce projet est simple : il s'agit de distribuer l'omble chevalier gratuitement aux femmes enceintes vivant dans les villages du Nunavik, habités majoritairement par des personnes d'origine inuite. Il est connu comme le Projet de distribution de l'omble chevalier (*Arctic Char Distribution Project* ou AC/DP dans son acronyme en anglais). Dans les années 2000, le projet a été mis à l'essai dans trois communautés du Nunavik. Au cours des années suivantes, on a tenté d'étendre le projet à tous les villages le long de la côte de la baie d'Hudson.

En septembre 2011, dix ans après sa première itération, la Régie régionale de la santé et des services sociaux du Nunavik (RRSSSN) a souhaité mettre en œuvre le projet de distribution d'omble chevalier pour toutes les femmes enceintes habitant les villages le long de la côte de la baie

d'Hudson. Théoriquement, le poisson est pêché par des pêcheurs locaux de cette zone, puis vendu légèrement en dessous du prix du marché à la Régie. Il est ensuite placé dans des glacières (de grande taille, difficiles à manipuler) afin d'éviter les risques de brûlures de congélation. Les glacières sont transportées en avion par la compagnie Air Inuit. Celle-ci transporte gratuitement les caisses de poisson aux villages environnants et, en théorie, ramène les caisses vides aux villages d'approvisionnement pour la somme modique de 25 dollars canadiens. Les maternités de ces villages doivent ensuite organiser la distribution du poisson frais aux femmes enceintes. Selon ce qui est prévu, toutes les femmes enceintes reçoivent un poisson de 24 pouces ou plus, une fois par semaine, à la maternité ou à leur domicile (dans les plus petits villages). Ce projet vise à réduire l'exposition aux contaminants environnementaux et à améliorer la sécurité alimentaire, deux problèmes de santé publique importants dans les communautés inuites du Nunavik.

Cependant, la mise en œuvre du projet est incomplète. Le poisson n'est pas régulièrement distribué (moins d'une fois par semaine) et toutes les zones géographiques ciblées ne sont pas couvertes. En outre, sa durabilité environnementale et financière est menacée.

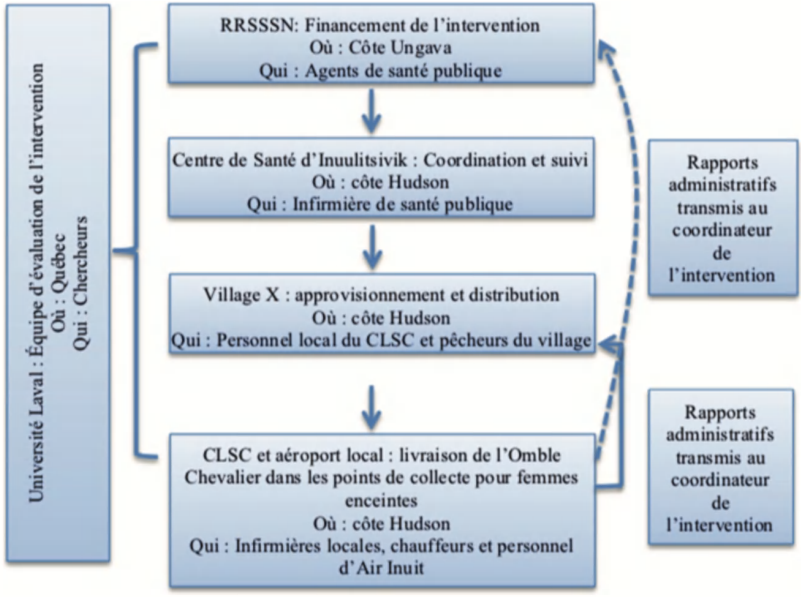


Figure 1. Schéma de la logique du projet et de ses composantes

Adaptation du cadre théorique au cas d'étude

Hasson (2010) avait constaté que les composantes du cadre de Caroll et ses collègues (2007) ne prenaient pas en compte le contexte socioéconomique, culturel et environnemental dans lequel un projet est implanté (Hasson 2010). Nous avons donc créé un thème supplémentaire, reflétant les différentes dimensions émises, qui était susceptible d'affecter la mise en œuvre du projet : « Valeurs associées à l'intervention » (voir section 4 ci-dessous). Les « valeurs » reflètent les différentes préconceptions naturellement associées au projet par les différentes catégories de participant-e-s. Elles se réfèrent au contexte macrosocial, économique, culturel et environnemental dans lequel l'intervention est mise en œuvre. Les participant-e-s parlent en effet de plusieurs types de valeurs associées à l'intervention :

- la valeur « de santé publique » (amélioration de l'état nutritionnel et sécurité alimentaire)
- la valeur « base communautaire » (intervention *bottom-up*, mise en œuvre au niveau communautaire)
- la valeur « traditionnelle » (préconiser la nourriture traditionnelle dans l'alimentation)
- la valeur « santé maternelle » (en référence à la population cible : les femmes enceintes)
- la valeur « lien avec les personnes âgées » (les générations âgées conseillant aux femmes enceintes de consommer du poisson)

Devis d'évaluation

Objectif de l'évaluation

Dans cette étude, nous conduisons une évaluation de la fidélité d'implantation du projet AC/DP. Plus précisément, nous examinons la manière dont les sept éléments de mesure de la fidélité d'implantation affectent l'atteinte des objectifs initiaux de ce projet, initié il y a 15 ans.

Déroulement

Nous évaluons l'implantation du projet à partir des données recueillies dans des documents sur le projet, comptes rendus de réunions, notes de terrain et entretiens qualitatifs avec les bénéficiaires et les personnes chargées d'exécuter le projet. Les thèmes émergents des données sont discutés à la lumière du cadre de fidélité d'implantation développé par Carroll et ses collègues en 2007.

Tableau 1. Récapitulatif du devis d'évaluation de l'implantation du Projet de distribution d'omble chevalier (Arctic Char) aux femmes enceintes du Nunavik

Sources de données	**Description**
Entretiens individuels	Guide d'entretien pour les bénéficiaires du projet – les femmes enceintes (à tout âge), explorant leur niveau d'acceptation du projet. Guide d'entretien pour les décideuses et décideurs et les personnes chargées de l'implantation, explorant les perceptions autour de la vision, du financement, de l'opérationnalité et de la viabilité du projet sur le long terme).
Documentation	Séjours sur le terrain (deuxième auteur) pour recueillir les documents.
Entretiens individuels	13 transcriptions verbatim – entretiens avec 14 femmes enceintes (un entretien s'est tenu avec deux femmes). 4 transcriptions verbatim – entretiens avec décideuses et décideurs et personnes chargées de l'implantation.
Documentation	7 documents – notes d'observation, rapports de réunion et documents de projet. 6 échanges de courriels
Transcription et thèmes	Après la transcription de toutes ces données, une approche standard de codage (déductif-inductif) a été réalisée. Les codes ont été examinés de façon collaborative par la première autrice et le deuxième auteur. Les thèmes potentiels ont été discutés à la lumière des objectifs de la recherche et ont été fixés par consensus.

Brève description des résultats de l'évaluation

Nos résultats indiquent que les femmes enceintes ont pleinement

accepté le projet en raison de sa pertinence culturelle. Cependant, on constate une implantation incomplète du projet qui se manifeste au niveau des sept dimensions :

1. **Impact attendu du projet** : absence de vision commune autour de l'impact attendu du projet : un projet visant la sécurité alimentaire pour les uns ou un projet visant la réduction de l'exposition aux contaminants pour les autres.
2. **Complexité inhérente à l'intervention** : logistique (poisson seulement disponible à certaines périodes de l'année, transport jusqu'aux bénéficiaires et maintien de la chaine du froid) et multiplicité des acteurs et actrices chargées de l'implantation (gestionnaires, coordinateurs, infirmières, etc.).
3. **Stratégies de communication et coordination** : absence de stratégies, difficultés de communication et de coordination parmi les personnes chargées de l'implantation du projet.
4. **Qualité de la prestation de l'intervention** : la fraicheur était parfois insatisfaisante, le poisson n'arrivant pas de façon régulière. En revanche, la taille du poisson semblait toujours satisfaire, et notamment le fait, pour les bénéficiaires, de pouvoir choisir une taille qui corresponde à leurs souhaits.
5. **Réceptivité au projet** : les bénéficiaires ont présenté des niveaux élevés de réceptivité – notamment parce qu'il s'agit, pour elles, d'une ressource gratuite. Le projet était également bien perçu par la plupart des professionnel-le-s de santé, conscient-e-s de l'utilité d'une telle intervention de santé publique dans ces communautés. En revanche, pour ce qui est de certains acteurs et actrices politiques, gestionnaires et personnels directement impliqués dans l'implantation, on pourrait parler d'un manque d'engagement et de vision de long terme : le leadership n'était pas toujours à la hauteur (quel que soit le niveau), les ressources humaines fluctuantes, et la structure de financement du projet instable.
6. **Liens de correspondance entre les caractéristiques du projet** : le poisson était de taille toujours satisfaisante, mais avec une couverture géographique partielle et des irrégularités dans la distribution (le poisson n'atteignant pas les bénéficiaires une fois par semaine pour des

raisons principalement logistiques).

7. **Valeurs associées** : intervention culturellement et socialement adaptée au contexte notamment parce qu'elle intègre des codes traditionnels (référence au partage, à la « nourriture locale », aux ancêtres, etc.). Intervention valable également du point de vue de la santé publique (nutriments et baisse de l'exposition aux contaminants). Potentiels risques du point de vue environnemental, lié une éventuelle surexploitation du poisson.

Tableau 2. Portrait récapitulatif des forces et faiblesses du projet

Forces du projet	**Faiblesses du projet**
• Apprécié et sollicité par les bénéficiaires (cf. n° 5) • Pertinence culturelle pour les communautés bénéficiaires (cf. n° 7) • Riche en nutriments (cf. n° 7) • Faible en contaminants (cf. n° 7) • Accepté par la plupart des professionnel-le-s de santé (cf. n° 5) • Coût limité (cf. n° 5)	• Complexité logistique, notamment des discontinuités dans la chaine de transport et de distribution pour s'assurer d'une distribution de poisson frais aux bénéficiaires du projet (cf n°6) • Défis de communication et de coordination parmi les personnes chargées de l'implantation du projet (cf. n° 3) • Absence de vision commune autour de l'impact attendu du projet (c.-à-d. projet visant la sécurité alimentaire pour les uns, et la réduction de l'exposition aux contaminants pour les autres) (cf. n° 1) • Durabilité financière insuffisante sur le long terme (cf. n° 5) • Changements fréquents des ressources humaines (notamment au sein des maternités et parmi les transporteurs) empêchant une implication sur le long terme des personnes responsables de l'implantation du projet • Caractère saisonnier du poisson et fraicheur irrégulière de la ressource liée à la manutention (cf. n° 2) • Potentielle vulnérabilité du poisson à la surexploitation (cf. n° 7)

Discussion des résultats

Nos résultats montrent que le projet a été soutenu et apprécié par de nombreux acteurs et actrices. Pourtant, ce travail met en évidence le fait que la complexité d'un projet peut entraver son implantation, notamment sur le plan de la communication et de la coordination.

La logistique (transport, expédition, stockage, distribution) nécessaire à la mise en œuvre du projet exige un niveau élevé de coordination entre les parties prenantes, comme les infirmières et les transporteurs communautaires, notamment parce que les poissons sont rapidement périssables.

Le manque d'engagement de certains acteurs et actrices a provoqué des difficultés dans l'implantation, de même que les changements fréquents de personnel dans la région du Nunavik. Outre la rotation du personnel d'Air Inuit et de la maternité, les coordonnateurs et coordonnatrices du projet ont souvent changé de poste. Les fréquentes rotations de personnel ont empêché le développement de réseaux coordonnés des personnes engagées sur le long terme, prêtes à surmonter les défis logistiques de la distribution des ressources dans une région éloignée. En tant que telle, la fiabilité du projet s'est dégradée dans le temps : les principales parties prenantes se révélaient souvent incapables de se prononcer sur la disponibilité du poisson. Ces défis rencontrés au Nunavik trouvent un écho en Afrique subsaharienne, région où l'instabilité du personnel est également commune. Plusieurs évaluations en Afrique subsaharienne font état de fréquents changements de personnel et de difficiles conditions de travail à la suite de la mise en place d'interventions de santé spécifiques au Sénégal et en Afrique du Sud par exemple (Rouleau et al. 2012; Van Damme, Kober, and Kegels 2008). Ces situations affectent fortement la fiabilité des projets.

Les problèmes structurels contribuent également à l'absence de communication entre les pivots du projet, en particulier les coordonnateurs et coordonnatrices de projets et les infirmières de la maternité, même lorsque travaillant à proximité. Des défis de communication ont également été identifiés dans l'évaluation de l'implantation d'une autre intervention en nutrition, le programme *Nutrition Nord Canada* (*Government of Canada* 2013). Des investigations sur ce phénomène dans d'autres projets ou

programmes de nutrition dans les régions éloignées du Canada devraient être poursuivies dans des recherches futures.

La complexité de l'intervention n'est probablement pas la seule responsable de l'implantation incomplète. En effet, comme l'évoque un décideur de la RRSSSN, la distribution d'omble chevalier aux femmes enceintes de cette région n'a jamais été considérée comme un « programme formel » piloté par la RRSSSN. L'intervention était plutôt perçue comme une simple « initiative », appuyée occasionnellement par la RRSSSN. Ceci expliquerait, toujours selon ce décideur, les difficultés d'implantation et le manque de visibilité sur le long terme.

En outre, l'absence d'une vision commune autour de l'impact prévu du projet a, selon nous, provoqué une dilution de la responsabilité entre les parties prenantes. L'engagement vers l'atteinte d'un objectif commun est pourtant la clé du succès du projet. Dans plusieurs cas, les acteurs et actrices de l'implantation n'étaient pas prêts à prendre l'initiative de trouver la source des problèmes de distribution. Pour résoudre les questions logistiques, le personnel des maternités se reposait souvent sur les coordonnateurs et coordonnatrices de projet et d'autres personnes en position d'autorité. Les fonctionnaires ont également démontré un engagement limité par leur connaissance restreinte du financement du projet et une restitution insuffisante des résultats des réunions aux institutions partenaires du projet. La responsabilité de toutes les parties prenantes par rapport aux bénéficiaires du projet a donc été remise en cause par un manque de leadership.

Nous estimons que le cadre développé par Carroll et ses collègues (2007) est un outil utile et valable pour guider les chercheuses et chercheurs et les décideuses et décideurs afin d'améliorer l'implantation de projets ou programmes d'intervention. Conçu comme une « recette de cuisine » dont il suffit de suivre chaque étape, le cadre s'est ainsi avéré utile pour évaluer de façon systématique les obstacles et facilitateurs à l'implantation du projet. Ceci est particulièrement pertinent dans le domaine de la santé des Autochtones en raison des niveaux de complexité élevés du contexte d'implantation. Il était nécessaire d'approfondir ces processus afin de mieux informer la prise de décision.

Rivages du village de Puvirnituq l'hiver, 2014. Crédits photo : Isabelle Veillette.

Cependant, l'absence de « valeurs » et d'éléments de contexte au sein du cadre de Carroll et ses collègues (2007) signale une vision technocratique de l'évaluation d'implantation de projets ou programmes qui vise à équilibrer facilitateurs et obstacles. Problématique majeure, le cadre ignore presque totalement le contexte humain et les liens entre les personnes impliquées. En particulier, il ne tient pas compte de la nécessité de disposer de personnes motivées pour la conception et l'implantation de projets ou politiques. Durlak et DuPre (2008) avaient pourtant montré à la même période que des facteurs liés à la capacité organisationnelle (motivation du personnel qui implante l'intervention, leadership, suivi et soutien des supérieurs hiérarchiques, etc.) étaient essentiels pour assurer une implantation complète (Chouinard et Cousins 2007). Ceci pose problème pour une application du cadre dans le contexte de la santé des Autochtones. L'établissement de liens entre les personnes y est primordial, car le collectif tient une place très importante. Dans le contexte autochtone, il s'avère donc nécessaire de le reconnaître dans toute évaluation réalisée.

Globalement, le projet AC/DP ne souffre pas d'un manque de soutien des parties prenantes, même si les objectifs spécifiques ont été souvent contestés. Le défi consiste plutôt à construire à partir de la popularité du projet afin de surmonter les obstacles logistiques et de communication.

Références clés

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J. et Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2(1), 40-48. https://doi.org/10.1186/1748-5908-2-40

L'article phare de Carroll et collègues détaillant le célèbre cadre théorique de la fidélité d'implantation : à lire et à relire.

Gautier, L., Pirkle, C. M., Furgal, C. et Lucas, M. (2016). Assessment of the implementation fidelity of the Arctic Char Distribution Project in Nunavik,

Quebec. *BMJ Global Health*, 1(3), e000093.
https://doi.org/10.1136/bmjgh-2016-000093

Notre article original de recherche, publié en anglais, dans lequel nous décrivons les différentes composantes du programme de distribution d'omble chevalier au Nunavik et les analysons à la lumière du cadre de la fidélité d'implantation.

Hasson, H. (2010). Systematic evaluation of implementation fidelity of complex interventions in health and social care. *Implementation Science*, 5(1), 67-75.
https://doi.org/10.1186/1748-5908-5-67

L'adaptation d'Hasson du cadre théorique de la fidélité d'implantation, tenant compte de manière plus convaincante du contexte sociopolitique et de l'environnement dans lequel une intervention complexe est implantée.

Références

Boucher, O., Jacobson, S. W., Plusquellec, P., Dewailly, E., Ayotte, P., Forget-Dubois, N., ... et Muckle, G. (2012). Prenatal methylmercury, postnatal lead exposure, and evidence of attention deficit/hyperactivity disorder among Inuit children in Arctic Quebec. *Environ Health Perspectives*, 120(10), 1456–1461.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J. et Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2(1), 40-48.
https://doi.org/10.1186/1748-5908-2-40

Chouinard, J. A. et Cousins, J. B. (2007). Culturally competent evaluation for Aboriginal communities: A review of the empirical literature. *Journal of Multidisciplinary Evaluation*, 4(8), 40–57.

Council of Canadian Academies. Expert Panel on the State of Knowledge of Food Security in Northern Canada. (2014). *Aboriginal Food Security in Northern Canada: An Assessment of the State of Knowledge.*
http://www.scienceadvice.ca/en/assessments/completed/food-security.aspx

Durlak, J. A. et DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and

the factors affecting implementation. *American Journal of Community Psychology*, *41*(3-4), 327–350.

Gautier, L., Pirkle, C. M., Furgal, C. et Lucas, M. (2016). Assessment of the implementation fidelity of the Arctic Char Distribution Project in Nunavik, Quebec. *BMJ Global Health*, *1*(3), e000093. https://doi.org/10.1136/bmjgh-2016-000093

Government of Canada. (2013). *Implementation Evaluation of the Nutrition North Canada Program.* Site de Aboriginal Affairs and Northern Development Canada. http://www.aadnc-aandc.gc.ca/eng/1395347953550/1395348287432

Hasson, H. (2010). Systematic evaluation of implementation fidelity of complex interventions in health and social care. *Implementation Science*, 5, 67-75. https://doi.org/10.1186/1748-5908-5-67

Jacobson, J. L., Muckle, G., Ayotte, P., Dewailly, É. et Jacobson, S. W. (2015). Relation of Prenatal Methylmercury Exposure from Environmental Sources to Childhood IQ. *Environmental Health Perspectives*, *123*(8), 827-833.

Lemire, M., Kwan, M., Laouan-Sidi, A. E., Muckle, G., Pirkle, C., Ayotte, P. et Dewailly, E. (2015). Local country food sources of methylmercury, selenium and omega-3 fatty acids in Nunavik, Northern Quebec. *Science of the Total Environment*, 509, 248–259.

Pirkle, C. M., Muckle, G. et Lemire, M. (2016). Managing mercury exposure in northern Canadian communities: State of the evidence and guidance for health professionals. *Canadian Medical Association Journal*, 188(14), 1015-1023.

Rouleau, D., Fournier, P., Philibert, A., Mbengue, B. et Dumont, D. (2012). The effects of midwives' job satisfaction on burnout, intention to quit and turnover: a longitudinal study in Senegal. *Human Resources for Health*, 10(1), 9–22.

Van Damme, W., Kober, K. et Kegels, G. (2008). Scaling-up antiretroviral treatment in Southern African countries with human resource shortage: How will health systems adapt?. *Social Science et Medecine*, 66(10), 2108–2121.

Van Oostdam, J., Donaldson, S. G., Feeley, M., Arnold, D., Ayotte, P., Bondy,

G., … et Loring, E. (2005). Human health implications of environmental contaminants in Arctic Canada : A review. *Science of the Total Environment*, 351, 165–246.

Résumé / Abstract / Résumé en inuktituk / Resumen

En 2011, la Régie régionale de la santé et des services sociaux du Nunavik a commencé à appuyer financièrement le Projet de Distribution d'Omble chevalier (*Arctic Char Distribution Project* ou AC/DP dans son acronyme en anglais) pour les femmes enceintes. Cette initiative favorisait la consommation d'une nourriture traditionnelle inuite – le poisson « Omble chevalier » – pour les femmes enceintes vivant dans les villages du Nunavik, une région du nord du Québec peuplée majoritairement de personnes d'origine inuite. Cette intervention visait à réduire l'exposition aux contaminants environnementaux et à améliorer l'état nutritionnel et la sécurité alimentaire. La mise en œuvre du projet a été évaluée sur la base des données collectées à partir de documents du projet, de notes de terrain et d'entretiens qualitatifs avec les bénéficiaires et les exécutants du projet. Les thèmes émergeant des données ont été discutés à la lumière du cadre de fidélité à la mise en œuvre développé par Carroll *et coll.* (2007). Les femmes enceintes ont pleinement adopté l'initiative pour sa pertinence culturelle. Toutefois, la mise en œuvre du projet était incomplète : elle ne couvrait pas toutes les zones géographiques prévues, et il y avait une incohérence récurrente dans l'approvisionnement et la distribution du poisson. En outre, l'initiative n'était pas financée de façon stable. Ce travail montre que la complexité du projet est susceptible d'entraver la réussite de la mise en œuvre, en particulier en termes de communication et de coordination.

***

In 2011, the Nunavik Regional Board of Health and Social Services began supporting the Arctic Char Distribution Project (AC/DP) for pregnant women. This initiative promoted consumption of a traditional Inuit food– the fish Arctic char – for pregnant women living in villages of Nunavik, an area in northern Quebec (Canada) inhabited predominantly by people of Inuit ethnicity. This intervention was intended to reduce exposure to environmental contaminants and improve nutriritional status and food security. The project's implementation was assessed based on data collected

from background documentation, field notes and qualitative interviews with project recipients and implementers. Themes emerging from the data were discussed in the light of the framework for implementation fidelity developed by Carroll et al in 2007. Pregnant women fully embraced the initiative for its cultural appropriateness. However, project implementation was incomplete because it did not cover all intended geographic areas, and there was a recurring inconsistency in the supply and distribution of the fish. In addition, the initiative has been inconsistently funded. This work highlights the extent to which project complexity can impede successful implementation, particularly in terms of communication and coordination.

***

Résumé en inuktituk

ᖃᐅᔨᓴᖅᑕᐅᓂᖓ ᑲᔪᓯᑎᑕᐅᑎᓴᖅᓯᒪᓕᖅᑐᑦ ᓄᑕᖅᓯᓴᓂᕐᒥᑦ ᐃᖃᓗᒃᑖᑎᑕᐅᖃᖅᑕᕐᓴᖅᓯᕐᓴᖅᓯᓕᕆᔨᖅᑕᑦ ᓄᓇᓕᓴᒥ

ᐱᐅᔪᓚᖅᐅᔪᓕᔪᖅ: -Γ, ᓄᓇᓕᓴᒥ ᐃᓗᔪᓕᓚᓂᖅᓯᒧᑦ ᑲᑎᒪᔨᑦ ᓴᓗᒥᖅᓯᕐᓚᐅᔪᓕᔪᑦ ᐃᖃᓗ2ᐲᑦ ᐊᓴᔪᑕᑦᑕᐅᓂᕐᖁᓕᑕ ᐱᓇᔪᐊᖅᑕᐅᓂᖓᓂᑦ GAC/DPH ᓄᑕᕐᓴᕐᓂᓄᓂᑦ. ᑕᒪᓇᓇ ᐱᑎᐊᖏᖓᔪᑎnzi ᐃᖃᓗ2ᐱᑐᖃᑦᑕᓂᕐᓴᐅᑲᕐᑐᑦ ᐊᖏᓂᑦ ᓄᑕᕐᓴᕐᓂᑦ ᓄᓇᓕᓂᒪᒥ, ᑯᐯᐢ ᑕᕐᓴᖏᓂ Gbᓇᑕᒥᕻ ᐃᓄᑐᐃᓄᖓᓂᑦ ᓄᓇᖏᖓᓂ. ᑕᒪᓇᓇ ᐱᑎᐊᖏᖓᐅᑎ ᑐᕋᔪᑎᐅᑲᓂᔪᓕᔪᖅ ᓂᐢᑯᓇᖏᑐᑕᓂᑦ ᐱᖏᓂᐅᒥᕐᑯᓂᑦᑐᓂᑦ ᐱᐅᓂᐅᒪᔨᑦᑕᐅᓂᑦ ᐱᖏᓂᐅᑎᓂᑦᑐᓗ ᐊᖏᓂᑦ ᓂᕿᑎᕐᑐᑦ.

ᐱᓂᐊᖏᖓᑦ: ᐱᓇᓂᐊᖏᕋᐅ2 ᑲᔪᓯᑎᑕᐅᓂᖓᓄ ᖃᐅᒥᕐᔨᖅᑕᐅᓂᓕᔪᖅ ᖃᐅᔨᓴᕐᑎᑎᓇᓂᑦ ᓄᐊᑕᐅᒪᔨᔪᓄᑦ, ᐊᓪᓚᓴᑯᑎᓚᐊᓇᓂᑦ ᖃᐅᔨᓴᕐᑎᑎᓄᑦ ᓄᓇᓚᓂᓂ ᐱᓇᐊᔪᓂᓕᑎᑦ ᐊᓪᓚᓗ ᐅᖃᖅᖃᑎᖃᖅᖃᓪᓚᓂᑦ ᐃᓚᐅᔪᓂᓇᑐᓂᑦ ᐊᑯᓂᓂᑯᓂᓕᐅᑎᓂ ᐱᕈᕐᓚᓂᐊᖏᓂᑦ ᑲᔪᓯᑎᑕᐅᓂᑦᑐᓗ. ᐃᓕᓂᓇᔪᑦ ᐱᑎᐊᖏᑎᑕᐅᓂᖓᓂᑦ ᖃᐅᔨᓴᕐᑎᑎᓇᓂᑦ ᓄᐃᑎᑎᑕᐅᓂᕐᔪᑦ ᐅᖃᐅᑎᖅᓯᓂᑦ ᐃᓗᓕᑎᑎᑕᐅᔪᑦ ᑲᔪᓯᑎᑕᐅᓂᕐᓴᐅᑎᓂᑦ WA5ᓕᔪᐃᓂᓴᖅ ᑭᐊᕐᑦ ᐊᖏᓗᓂ ᓱᓄᓂ -Γ.

ᓄᐃᑎᑕᐅᓂᖓᑦ: ᓄᑕᕐᓴᕐᓂᑦ ᐊᖏᓂᑦ ᐃᓚᐅᑦᔪᐊᑐᐃᓇᓂᑦ ᐱᑎᐊᖏᖅᓄᓱᓇᔪᐊᕐᓴᐅᑎᓂᑦ ᐃᓗᕐᔨᑎᓇᓂᕐᓯᐊᒪᓚᖅᑕᐅᔪᑦ. ᑌᒪᓗᓗᐊᕐᓴᖅᓗᔪ, ᐱᓇᔪᐊᑕᕐᓴᖅᐅᔪᓕᔪᖅ ᐅᓪᓚᓕᔪᖅᑕᐅᓂᖓᓂᑦ ᐅᐱᔪᓇᔪᓚᓕᒪᑦ, ᐃᖃᓗᒃᑖᑎᑕᐅᓂᖓᓂᑦ ᐅᐱᔪᓇᔪᓚᓕᒪᑦ ᐊᖏᓂᑦ ᐃᓗᕐᔨᑎᓇᓂᑦ ᐊᑯᓂᓂᑯᓂᓕᐅᑎᓂ ᐱᕈᕐᓚᓂᐊᖏᓂᑦ ᐅᐱᔪᓇᔪᓚᓕᒪᑦ ᐊᖏᓂᑦ ᐃᓗᕐᔨᑎᓇᓂᑦ ᐱᑎᐊᖏᖅᓄᓱᓇᔪᐊᕐᓴᐅᑎᓂᑦ ᐱᖏᓂᐅᒥᕐᑯᓂᑦ ᐅᐱᔪᓇᔪᓚᓕᒪᑦ ᐊᖏᓂᑦ ᐃᓗᕐᔨᑎᓇᓂᑦ ᐱᑎᐊᖏᖅᓄᓱᓇᔪᐊᕐᓴᐅᑎᓂᑦ ᐱᖏᓂᐅᒥᕐᑯ

***

En 2011, la Junta Regional de Salud y Servicios Sociales de Nunavik comenzó a proporcionar apoyo financiero al Arctic Char Distribution Project (AC/DP) para mujeres embarazadas. Esta iniciativa promovió el consumo de alimentos tradicionales de los inuit, el salmón del ártico, para las mujeres embarazadas que viven en las aldeas de Nunavik, una región del norte de Quebec poblada principalmente por personas de origen inuit. Esta intervención tenía por objeto reducir la exposición a los contaminantes ambientales y mejorar el estado nutricional y la seguridad alimentaria. La ejecución del proyecto se evaluó sobre la base de los datos recogidos de los documentos del proyecto, las notas sobre el terreno y las entrevistas cualitativas con los beneficiarios y ejecutores del proyecto. Los temas que surgieron de los datos fueron discutidos a la luz del marco de fidelidad de implementación desarrollado por Carroll et al. (2007). Las mujeres embarazadas han adoptado plenamente la iniciativa por su relevancia cultural. Sin embargo, la ejecución del proyecto fue incompleta: no abarcó todas las zonas geográficas previstas y había una incoherencia recurrente en el suministro y la distribución de pescado. Además, la iniciativa no se financió sobre una base estable. Este trabajo muestra que la complejidad del proyecto puede dificultar una ejecución exitosa, especialmente en términos de comunicación y coordinación.

***

Lara Gautier est candidate au doctorat en santé publique, option santé mondiale, à l'École de Santé publique de l'Université de Montréal. Formée en sciences politiques et en économie de la santé, Lara s'intéresse aux enjeux de gouvernance et de pouvoir dans les interventions de santé mondiale, en utilisant une approche interdisciplinaire d'analyse des processus politiques, socioculturels et organisationnels. Ses terrains de recherche couvrent notamment l'Afrique de l'Ouest francophone et les populations vulnérables au Québec et en France. Elle est aussi impliquée dans le milieu associatif, apportant un soutien social aux migrants à Montréal et à Paris. Lara a publié plus d'une quinzaine d'articles scientifiques et chapitres d'ouvrage. Elle est titulaire d'une bourse de recherche au doctorat du FRQ Société et Culture, et enseigne la santé mondiale, l'évaluation d'implantation et les méthodes qualitatives et mixtes au Québec, au Mali et dans plusieurs pays d'Europe.

Catherine Pirkle est professeure adjointe à l'Office of Public Health

Studies de l'Université d'Hawaï'i à Mānoa. Elle est l'une des chercheuses principales de l'équipe d'évaluation de *Healthy Hawaii Initiative*. Cette initiative de l'État d'Hawaï, démarrée en 2000,a mis en œuvre plusieurs interventions de santé publique importantes,visant notamment à promouvoir l'activité physique, l'alimentation saine, le renforcement des systèmes de santé. Financée par le Département de santé d'Hawaï et les Centers for Disease Control des États-Unis, l'équipe de Pirkle et de ses collègues est chargée d'évaluer les principales interventions de *Healthy Hawaii Initiative*. Les domaines de recherche de Dre Pirkle sont la santé mondiale, la santé des femmes, et le vieillissement. Outre son travail pour l'équipe d'évaluation de *Healthy Hawaii Initiative*, elle est investigatrice principale ou co-investigatrice de plusieurs projets financés par les National Institutes of Health des États-Unis et les Instituts de recherche en santé du Canada.

Christopher Furgal est professeur agrégé dans le programme d'études et de sciences environnementales autochtones à l'Université Trent. Il détient une formation en sciences naturelles, sciences sociales et de la santé. Il est nommé conjointement à l'École d'Études Autochtones Chanie Wenjacket l'École Trent pour l'Environnement. Il est Directeur du Groupe de Recherche sur la Santé, l'Environnement et les Communautés Autochtones, et Directeur du Centre Nasivvik pour la Santé des Inuits et les Environnements Changeants. Dr Furgal s'intéresse à la santé environnementale autochtone et à l'interaction entre la science et les connaissances autochtones, notamment dans l'Arctique canadien. Il travaille depuis 25 ans avec des communautés autochtones (communautés inuites en particulier) et des organisations sur des questions de santé environnementale y compris le changement climatique et la santé, la sécurité alimentaire, l'exposition aux contaminants et la perception des risques. Il est co-auteur ou auteur principal de plusieurs évaluations nationales et internationales en santé environnementale.

Michel Lucas est professeur agrégé au Département de médecine sociale et préventive de la Faculté de médecine de l'Université Laval, chercheur au CHU de Québec-Université Laval et chercheur invité à la Harvard T.H. Chan School of Public Health. Le domaine de recherche du Dr Lucas est la qualité alimentaire, plus spécifiquement les indices multidimensionnels de la qualité alimentaire. Ses activités de recherche en santé mondiale s'intéressent à l'effet global de l'alimentation dans la santé des populations à partir de données d'études populationnelles variées

(enfants, adultes canadiens, Autochtones du Québec, Polynésiens, etc.). Comme chercheur principal désigné, il dirige le vaste projet d'intervention VisezEau® regroupant 14 chercheuses et chercheurs de différentes universités du Québec, de l'Ontario et des États-Unis. Dr Lucas a reçu plusieurs subventions, bourses et prix en reconnaissance de ses travaux. Membre de plusieurs comités éditoriaux et réviseur pour de prestigieuses revues scientifiques, il a réalisé plus de 300 publications et présentations.

Citation

Lara Gautier, Catherine M. Pirkle, Christopher Furgal et Michel Lucas. (2019). L'évaluation de la fidélité d'implantation. Un projet de distribution d'Omble chevalier aux femmes enceintes du Nunavik. In *Évaluation des interventions de santé mondiale. Méthodes avancées*. Sous la direction de Valéry Ridde et Christian Dagenais, pp. 383-401. Québec : Éditions science et bien commun et Marseille : IRD Éditions.

16. L'évaluation de la fidélité et de l'adaptation

Un exemple de mise en œuvre des interventions en santé mondiale

DENNIS PÉREZ, MARTA CASTRO ET PIERRE LEFÈVRE

L'évaluation de la fidélité et de l'adaptation

Définition de la méthode

Il s'agit d'une analyse holistique visant à étudier la manière dont les acteurs et actrices mettent en œuvre les interventions telles que planifiées ou la manière dont ils et elles introduisent des changements. Cela rend la mise en œuvre de qualité et efficace.

Forces de la méthode

- Permet d'identifier les composantes essentielles indispensables à mettre en œuvre pour que l'intervention produise ses effets attendus.
- Permet d'obtenir un équilibre adéquat entre fidélité et adaptation dans le cas d'interventions adaptables.
- Permet d'identifier des stratégies de facilitation pour faire face aux adaptations inadéquates.
- Aide à distinguer entre l'échec de l'intervention (manque d'efficacité) et échec de la mise en œuvre.

Défis de la méthode

- Quand une méthode standardisée pour évaluer la fidélité n'est pas disponible et que les mesures proposées sont limitées et souvent spécifiques aux programmes évalués.
- Manque de consensus sur la manière d'opérationnaliser les adaptations.

Ce chapitre concerne l'évaluation de la fidélité et des adaptations dans la mise en œuvre des interventions de santé mondiale. Il s'intéresse particulièrement aux interventions adaptables. En effet, la plupart des interventions en santé mondiale sont susceptibles de subir des adaptations, car elles sont complexes, abordent des problèmes de grande envergure, impliquent différents niveaux organisationnels et ciblent des comportements collectifs. Le contrôle de leur mise en œuvre par les

concepteurs et conceptrices est souvent impossible, voire même non souhaitable. De plus, elles dépendent du contexte, car certains programmes peuvent bien fonctionner dans certains contextes et échouer dans d'autres.

La fidélité avec laquelle est mise en œuvre une intervention peut affecter son efficacité. Par conséquent, atteindre un haut niveau de fidélité est une préoccupation majeure quand on passe de la recherche portant sur l'efficacité d'une intervention dans des conditions idéales à sa mise en œuvre dans le monde réel. Lorsqu'une intervention est mise en pratique, la survenue d'adaptations est plus que probable. Un contrôle total du processus de mise en œuvre sera souvent impossible, ni même souhaitable, et de nombreuses interventions devront être adaptées.

Pareillement à la fidélité, les adaptations peuvent avoir une incidence positive ou négative sur l'efficacité d'une intervention de santé mondiale. De ce fait, l'évaluation des deux aspects est importante.

La mise en œuvre est un ensemble spécifique de processus et d'activités, conçus pour mettre en pratique une intervention dont les dimensions sont connues (Fixsen *et al.* 2005). Pour l'évaluer, il faut considérer que les résultats de la mise en œuvre sont conceptuellement et empiriquement distincts de ceux utilisés pour évaluer l'efficacité de l'intervention (Proctor *et al.* 2011; Peters *et al.* 2013). La distinction entre les effets de la mise en œuvre et les effets d'une intervention est cruciale. Lors d'un échec d'une intervention, cette différenciation permet de déterminer s'il est dû à l'inefficacité de l'intervention elle-même (échec de l'intervention), ou si sa mise en œuvre s'est faite incorrectement (échec de la mise en œuvre) (Proctor *et al.* 2011).

La fidélité, ou le degré avec lequel une intervention est mise en œuvre comme prévu par ses concepteurs (Dane et Schneider 1998; Dusenbury *et al.* 2003; Schwarz *et al.*, 2014; Sundell *et al.* 2015), est un résultat du processus de mise en œuvre (Proctor *et al.* 2011). Il est particulièrement destiné à assurer que l'intervention maintienne ses effets prévus (Dane et Schneider 1998; Backer 2002; Dusenbury *et al.* 2003; Schwarz *et al.* 2014; Sundell *et al.* 2015). Cette notion est proche de termes tels que l'*intégrité* (Dane et Schneider 1998, Dusenbury et al. 2003), le *fondement* (Fixsen *et al.* 2005, Rebchook *et al.* 2006, Schwarz *et al.* 2014, Gagliardi *et al.* 2014, Sundell *et al.* 2015) ou encore les *composantes essentielles* de l'intervention (Carroll *et al.* 2007), c'est-à-dire celles qui doivent être nécessairement mises en œuvre si l'on souhaite que l'intervention produise ses effets.

L'adaptation, au contraire, est la modification du modèle d'origine d'une intervention (Backer 2002; Schwarz *et al.* 2014; Sundell *et al.* 2015). Dans la théorie de la diffusion des innovations (Rogers 2003), l'adaptation est comprise comme le résultat de processus cognitifs qui se produisent lorsque les acteurs et actrices ont du mal à donner un sens à une intervention au cours de sa mise en œuvre. Dans cette compréhension du terme, ce type spécifique d'adaptation, qui dépasse la simple adaptation aux caractéristiques locales, est très souvent appelé la *réinvention*.

Les adaptations peuvent être délibérées ou accidentelles et comprennent plusieurs modalités : l'ajout de nouvelles composantes, la suppression ou la modification radicale d'une composante de l'intervention de telle sorte qu'elle ne ressemble plus à l'original ou des modifications mineures ou majeures apportées à une composante (Backer 2002; Rebchook *et al.* 2006).

La fidélité avec laquelle une intervention est mise en œuvre affecte son efficacité (Dane et Schneider 1998; Backer 2002; Dusenbury *et al.* 2003; Carroll *et al.* 2007; Schwarz *et al.* 2014; Sundell *et al.* 2015). Par conséquent, l'atteinte d'un haut niveau de fidélité a été la principale préoccupation de nombreux chercheurs et chercheuses aux prises avec les difficultés à passer des études portant sur l'efficacité des interventions dans des conditions idéales à leur mise en œuvre dans le monde réel (Bellg *et al.* 2004; Borrelli 2011). Cependant, dans la pratique, l'adaptation des interventions est la règle plutôt que l'exception (Sundell *et al.* 2015). De plus, certaines interventions doivent nécessairement être adaptées au cours de leur mise en œuvre (Backer 2002; Schwarz *et al.* 2014; Sundell *et al.* 2015). C'est le cas des *interventions adaptables*.

Les interventions adaptables sont celles où les acteurs et actrices de la mise en œuvre sont autorisés, ou même encouragés, à apporter des modifications au modèle initial. Cette définition comprend le type d'interventions adaptables telles que définies par Collins (*et al.* 2004) pour lesquelles des changements préétablis et anticipés par les concepteurs et conceptrices de l'intervention sont autorisés. Pour ce type particulier d'intervention, la fidélité est importante pour s'assurer que les adaptations prédéfinies se sont produites comme prévu. Cependant, les interventions adaptables comprennent également celles où des changements non prédéfinis sont induits par les acteurs et actrices de première ligne. Au cours des quinze dernières années, cinq dimensions ont été rapportées pour

mesurer la fidélité : i) l'adhérence – la mise en œuvre du programme comme prévu, ii) la dose – la fréquence et la durée de l'exposition au programme, iii) la qualité de la prestation – la manière dont le programme est délivré, iv) la réactivité des participant-e-s – le degré auquel les participant-e-s sont engagés et v) la différenciation du programme – les caractéristiques critiques qui distinguent le programme (Dane et Schneider 1998; Backer 2002; Dusenbury *et al.* 2003; Carroll *et al.* 2007; Schwarz *et al.* 2014; Sundell *et al.* 2015). Cependant, peu de recherches ou de conseils pratiques portent sur la façon d'adapter une intervention tout en maintenant ses composantes et mécanismes essentiels à son efficacité (Schwarz *et al.* 2014).

Sur la base d'une revue systématique et critique des conceptualisations existantes de la fidélité de la mise en œuvre, Carroll et ses collègues (2007) ont proposé un cadre conceptuel pour donner du sens au concept de fidélité, mais aussi pour clarifier et expliquer la fonction de chacune des cinq dimensions classiques et leurs liens. Dans leur cadre conceptuel, ils et elle ont également inclus deux éléments supplémentaires : la complexité des interventions et les stratégies de facilitation.

Selon ce cadre, « la mesure de la fidélité de la mise en œuvre est la mesure de l'adhérence » (Carroll *et al.*, en ligne) c'est-à-dire à quel point les responsables de la mise en œuvre de l'intervention adhèrent à l'intervention telle que décrite par ses concepteurs et conceptrices. L'adhérence comprend le contenu, la fréquence, la durée et la couverture (la dose). Le niveau de réalisation du contenu prévu ou de la fréquence souhaitée est le degré de fidélité de la mise en œuvre atteint pour cette intervention. Ce niveau peut être influencé par d'autres variables : la complexité de l'intervention, les stratégies de facilitation, la qualité de la mise en œuvre et la réactivité des participant-e-s (Figure 1). La relation entre une intervention et son impact (représentée par une ligne discontinue dans la figure) est externe à la fidélité de la mise en œuvre, mais le degré de fidélité atteint peut affecter cette relation. L'évaluation des résultats doit prendre en considération le contenu de l'intervention en déterminant les exigences minimales (composantes essentielles) nécessaires pour maintenir un niveau de fidélité élevé. Mais tant que ces composantes ne sont pas connues, la fidélité à tous les éléments de l'intervention demeure nécessaire (Carroll *et al.* 2007).

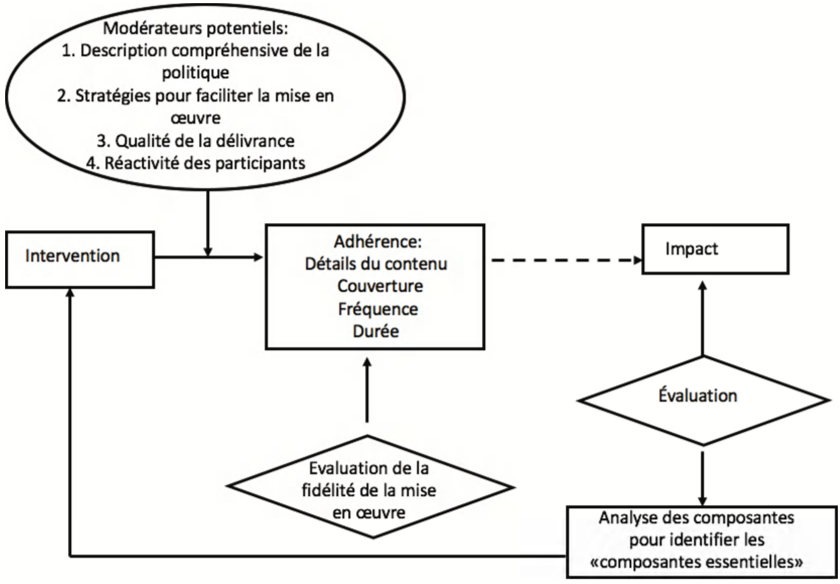


Figure 1. Modèle conceptuel pour l'évaluation de la fidélité de mise en œuvre proposé par Carroll et al. 2007

Nous soutenons l'idée que fidélité et adaptation coexistent (Rogers 2003). De notre point de vue, les adaptations peuvent conduire à améliorer ou à menacer la théorie du changement sous-jacente à l'intervention (Hernandez et Hodges 2003) et donc avoir un impact positif ou négatif sur son efficacité. Dès lors, atteindre un équilibre approprié entre fidélité et adaptation lors de la mise en œuvre est nécessaire. L'une des façons d'aborder cette question est de rechercher systématiquement les aspects du programme qui ont subi des adaptations lorsque l'on évalue sa mise en œuvre. Fidélité et adaptation ne sont pas nécessairement mutuellement exclusives. Les programmes peuvent incorporer fidélité et adaptations. Dès lors, l'évaluation doit examiner attentivement les deux aspects. Pour ce faire, les chercheuses et chercheurs ont besoin a) d'une description claire et complète de l'intervention, y compris de son contenu (quoi faire, pour combien de temps et par qui), comment mettre l'intervention en pratique et ses principes de fonctionnement, c'est-à-dire la théorie du changement sous-jacente à l'intervention et b) les mesures de fidélité et d'adaptation

par rapport aux caractéristiques de l'intervention (complexité, caractère innovant), les outils d'évaluation et les sources de vérification.

Dans le reste de ce chapitre, nous fournissons un exemple d'application d'une évaluation de la fidélité d'une intervention adaptable dans le contexte particulier de Cuba et nous en tirons un certain nombre de leçons.

Mise en contexte

La dengue est une maladie à transmission vectorielle qui est principalement transmise par un moustique femelle du genre *Aedes aegypti* infecté par l'un des quatre virus de la dengue (Stoddard *et al.* 2013). Un épisode simple de la maladie peut évoluer vers une dengue hémorragique sévère ou mortelle (Guzmán *et al.* 2006). Toutes les méthodes de contrôle actuellement disponibles ciblent le vecteur. Elles sont presque impossibles à mettre en œuvre sans l'acceptation de la communauté ou sa participation active (Parks et Lloyd 2004; Toledo *et al.* 2008).

Les stratégies de renforcement des capacités à agir ont été rapportées comme efficaces pour la prévention et le contrôle de la dengue à l'échelle communautaire à Cuba (Toledo *et al.* 2007; Sánchez *et al.*, 2009; Vanlerberghe *et al.* 2009; Castro *et al.* 2012) et ailleurs (Kay et Nam 2005; Kay *et al.* 2010). Le renforcement des capacités à agir est un processus par lequel les individus, les groupes et les communautés se voient dotés des capacités nécessaires pour renforcer leur pouvoir à agir sur les décisions qui influencent leur vie (Rifkin 1996); dans notre cas, ce sont les déterminants des épidémies de dengue. Pourtant, ces stratégies demeurent controversées en raison des problèmes de transférabilité et de mise à l'échelle. De plus, les descriptions des processus de leur mise en œuvre sont rares dans la littérature (Rifkin 2014).

Les preuves scientifiques qui lient directement la participation de la communauté aux impacts sur la santé sont généralement considérées comme limitées et insuffisantes. Lorsque de tels liens causaux sont trouvés, ils sont considérés comme spécifiques au contexte et au processus (contenu). Mais la mise en œuvre et le rôle du contexte sont ignorés ou mal décrits et analysés dans la littérature (Rifkin 2014). Ainsi, nous avons réalisé un ensemble d'études en vue de fournir une description détaillée et solide des processus de mise en œuvre d'une stratégie de renforcement des

capacités à agir pour la prévention et le contrôle de la dengue à Cuba. Nous abordons les défis posés par la fidélité et l'adaptation rencontrés lors de sa reproduction et mise en œuvre à plus grande échelle.

Une stratégie de renforcement de la capacité à agir fondée sur les données probantes pour la prévention de la dengue, jumelée au Programme cubain de contrôle d'*Aedes aegypti* (AaCP), a été conçue puis répliquée à une échelle plus large (Pérez *et al.* 2010; Castro *et al.* 2012). L'AaCP, créé en 1981, a une couverture nationale. Il est structuré selon les niveaux organisationnels du système national de santé (national, provincial, municipal et primaire). Il est verticalement structuré et géré principalement au niveau national bien que, dans une certaine mesure, une prise de décision décentralisée soit possible en fonction des caractéristiques locales et de la situation épidémiologique. Pendant des années, des milliers de travailleurs de terrain ont mené des activités standardisées telles que la surveillance entomologique, la réduction des sources larvaires et le contrôle des moustiques adultes, l'éducation à la santé et l'application de la législation sur le contrôle des moustiques (c'est-à-dire l'utilisation d'amendes lorsque les lieux de reproduction de moustiques sont retrouvés dans les ménages). Dans l'approche de l'AaCP, le rôle de la population demeure cependant limité et passif (Kourí *et al.* 1989).

Description de la stratégie innovante de renforcement des capacités à agir pour la prévention de la dengue

La stratégie a été développée par des chercheuses et chercheurs de l'Institut de médecine tropicale Pedro Kourí à La Havane. Elle est basée sur le renforcement des capacités à agir et les théories d'éducation populaire (Freire 1972). Elle vise la participation communautaire dans la prise de décision pour les activités de lutte antivectorielle. Conçue comme une composante horizontale à intégrer dans le programme de contrôle vertical existant, elle comprend quatre composantes principales : l'organisation et la gestion, le renforcement des capacités, le travail communautaire et la surveillance. Pour chaque composante, les activités (c'est-à-dire les sous-composantes) qui devaient être mises en œuvre ont été définies (Table 1).

La stratégie a été reproduite entre octobre 2004 et décembre 2007 dans 16 communautés (circonscriptions) sélectionnées au hasard dans trois

Conseils Populaires de la municipalité de La Lisa (Castro *et al.* 2012). La circonscription est le niveau le plus décentralisé du gouvernement local (environ 1 000 habitants). Le Conseil Populaire est une structure gouvernementale intermédiaire entre le niveau municipal et la circonscription. Les circonscriptions et les Conseils Populaires étaient hétérogènes sur le plan de la composition sociodémographique de la population, d'expériences antérieures avec la participation communautaire, de caractéristiques du leadership, de ressources et de dynamique de la gouvernance locale.

Tableau 1. La stratégie de renforcement de la capacité à agir pour le contrôle d'Aedes aegypti par composantes et sous-composantes. Municipalité de La Lisa, La Havane, 2004-2007

Composantes	Définition opérationnelle	Sous-composantes
1. Renforcement des capacités	Développement des connaissances, des capacités et des valeurs et pratiques associées nécessaires aux membres de la communauté pour conduire la stratégie.	1.1 Diagnostic, travail de groupe et participation
		1.2 Surveillance des risques et comportements
		1.3 Plans d'action et stratégie de communication
		1.4 Évaluation participative
2. Organisation et gestion	La manière dont les acteurs et actrices impliqués dans le contrôle de la dengue se définissent, s'organisent, s'engagent et se fixent des rôles, identifient les ressources et prennent des décisions	2.1 Présence de groupes de travail communautaires (GTCs) conduisant la stratégie
		2.2 Présence du personnel du programme de contrôle vectoriel dans les CTCs
		2.3 Identification des ressources communautaires
		2.4 Mobilisation des ressources extérieures
3. Travail communautaire	Cycle répétitif d'actions développé par un groupe issu de la communauté pour modifier les conditions qui accroissent la probabilité de transmission de la dengue	3.1 Cartographie des risques
		3.2 Diagnostique des problèmes
		3.3 Plan d'action
		3.4 Actions exécutées
		3.5 Stratégie de communication
		3.6 Élaboration des matériaux de communication
		3.7 Suivi et évaluation

4. Surveillance	Collecte et analyse, systématique et en temps opportun, des données portant sur les risques de transmission de la dengue et les comportements associés pour l'action	4.1 Identification des risques environnementaux
		4.2 Identification des risques domiciliaires
		4.3 Identification des comportements associés

Protocole de recherche et cadre d'analyse

Une évaluation rétrospective de la mise en œuvre a été effectuée dans le cadre de l'évaluation finale de la reproduction de la stratégie. Ses objectifs étaient les suivants : 1) évaluer l'équilibre fidélité-adaptation dans la reproduction de la stratégie ; 2) comprendre les difficultés et les obstacles rencontrés lors de la mise en œuvre dans le but d'améliorer le processus ; et 3) faire progresser les connaissances théoriques et pratiques quant à la conduite de la recherche sur la fidélité dans le cas d'interventions adaptatives.

Pour atteindre nos objectifs, nous avons mené une évaluation de la fidélité en utilisant une méthode mixte de quantification de données qualitatives (Creswell et Clarck 2011).

Par définition, une stratégie de renforcement des capacités à agir se doit d'être modifiée et révisée en fonction de son interaction avec le cadre institutionnel et les contextes dans lesquels elle est mise en œuvre (Castro *et al.* 2008). Compte tenu du caractère adaptatif de la stratégie, nous avions anticipé que les dimensions classiques de la fidélité auraient une utilité limitée pour évaluer les adaptations. En effet, la promotion de la participation communautaire à la prise de décision implique, par définition, une certaine incertitude quant à ce qui sera planifié et/ou atteint.

Nous avons donc choisi d'évaluer la fidélité et l'adaptation dans la mise en œuvre en utilisant la méthodologie de Rebchoock (*et al.* 2006). Selon ces auteurs, trois types d'adaptation à la conception initiale d'une intervention peuvent être identifiés : (1) l'omission ou la modification d'une composante, si bien que le programme n'est plus mis en œuvre comme prévu, (2) une composante mise en œuvre avec des modifications mineures ou majeures, tout en respectant le but initial ou (3) des activités supplémentaires ou

des composantes ajoutées. En théorie, les implications de ces trois types d'adaptations sur la fidélité sont différentes. Lorsqu'une composante est supprimée, la fidélité est menacée. Dans le second cas, selon ce que les modifications entraînent, la fidélité peut être affectée ou non. Enfin, en ajoutant quelque chose de nouveau, la fidélité peut être facilement maintenue.

Application de la méthode et des procédures pour évaluer la fidélité et l'adaptation

Une évaluation en trois étapes a été menée (Pérez *et al.* 2010). La première étape était une évaluation individuelle par les trois conceptrices de la stratégie impliqués en tant que facilitatrices de sa mise en œuvre. Elle s'est basée sur l'analyse des procédures du projet, des procès-verbaux des ateliers réalisés et l'examen des documents remplis par les acteurs et actrices de la mise en œuvre. La deuxième étape était une discussion sur les conclusions obtenues avec un groupe élargi d'expert-e-s (les trois impliqués dans la première étape et trois nouveaux). Si tous et toutes s'accordaient sur le fait qu'une composante ou une sous-composante avait été mise en œuvre comme prévu dans une circonscription donnée, elle était classée comme mise en œuvre. En cas de consensus sur le fait qu'une composante ou une sous-composante n'était pas mise en œuvre, elle était classée comme non mise en œuvre. Finalement, si l'un-e des expert-e-s estimait qu'une composante ou une sous-composante avait été modifiée, elle était classée comme modifiée. Les activités ajoutées ont également été identifiées. La troisième étape était une évaluation progressivement affinée de manière consensuelle à l'issue de discussions avec les responsables de la mise en œuvre à travers un atelier d'évaluation participatif.

Les données qualitatives obtenues à partir de l'évaluation en trois étapes (c'est-à-dire les descripteurs de la mise en œuvre des composantes/sous-composantes de la stratégie pour chaque circonscription) ont été transformées en données quantitatives (Creswell et Clarck 2011). Les fréquences des sous-composantes non mises en œuvre, modifiées et mises en œuvre ont été tabulées pour toutes les circonscriptions. La moyenne a été calculée pour les quatre composantes. Pour explorer les relations entre les principales composantes, un score de 0, 1 et 2 a été attribué

respectivement aux composantes omises, modifiées et mises en œuvre de la stratégie. Leurs valeurs ont été additionnées pour chaque circonscription. Par la suite, nous avons calculé les coefficients de corrélation Spearman Rank entre les composantes en utilisant le logiciel SPSS V.15.

De plus, des entrevues semi-structurées ont été menées avec treize informateurs et informatrices clés pour identifier les explications possibles aux variations observées dans la mise en œuvre. Un échantillonnage intentionnel a été utilisé pour la sélection des personnes enquêtées, afin d'identifier celles ayant un rôle important dans la mise en œuvre, à différents niveaux et fonctions. Le guide d'entrevue a exploré la compréhension des personnes interrogées en ce qui concerne les composantes/sous-composantes de la stratégie, les difficultés qui ont surgi au cours de la mise en œuvre et leurs causes perçues. Les entretiens ont été enregistrés, retranscrits et analysés avec l'appui du logiciel QSR Nvivo 7. L'analyse a été principalement déductive, permettant cependant l'émergence de thèmes supplémentaires (par exemple les difficultés).

Principaux résultats

La surveillance a été la composante la plus mise en œuvre, suivie du renforcement des capacités. Le travail communautaire et l'organisation et la gestion ont été les composantes les moins mises en œuvre ou modifiées (Table 2) (Pérez *et al.* 2010). Même pour les composantes les plus mises en œuvre, des sous-composantes ne l'ont pas été. Plus les composantes et les sous-composantes étaient novatrices, moins elles ont été mises en œuvre, par exemple celles qui impliquaient des activités différentes ou perturbatrices des pratiques courantes du contrôle vectoriel et qui ne pouvaient pas être facilement réalisées en faisant appel à l'expertise, aux expériences antérieures ou au savoir-faire des acteurs et actrices de la mise en œuvre.

COMPOSANTES DE LA STRATÉGIE	Conseil Populaire 1				
	1	2	3	4	5
SURVEILLANCE					
Identification des risques environnementaux	I	I	I	I	I
Identification des risques domiciliaires	I	I	I	I	I
Identification des comportements associés	I	I	N	I	N
RENFORCEMENT DES CAPACITÉS					
Diagnostique, travail de groupe & participation	I	I	I	I	I
Surveillance des risques et comportements	I	I	I	I	I
Planification et stratégie de communication	I	I	M	I	M
Évaluation participative	I	I	I	I	N
Ateliers "auto-organisés"	A				
TRAVAIL COMMUNAUTAIRE					
Cartographie des risques	I	I	I	I	M
Diagnostique des problèmes	I	I	I	M	M
Plan d'action	I	I	N	N	N
Actions exécutées	I	I	N	N	N
Stratégie de communication	I	I	M	M	M
Élaboration des matériaux de communication	I	N	N	N	N
Suivi et évaluation	I	I	I	I	I
Évaluation des points forts de la communauté	A	A	A		
ORGANIZATION & GESTION					
Présence des GTCs conduisant la stratégie	I	I	M	M	N
Personel *Aa*CP dans les GTC	I	M	M	M	N
Identification des ressources communautaires	I	I	N	N	N
Mobilisation des ressources extérieures	I	I	N	N	N
Lien avec l'école primaire établi	A			A	
Implication d'autres acteurs	A	A	A	A	A

Table 2a. Composantes et sous-composantes de la stratégie mises en œuvre, modifiées, non mises en œuvre ou ajoutées par le Conseil Populaire 1 et les circonscriptions. Municipalité de La Lisa, La Havane, 2004-2007

COMPOSANTES DE LA STRATÉGIE	Conseil Populaire 2					
	6	7	8	9	10	11
SURVEILLANCE						
Identification des risques environnementaux	I	I	I	I	N	N
Identification des risques domiciliaires	I	I	I	I	I	I
Identification des comportements associés	I	N	N	N	N	N
RENFORCEMENT DES CAPACITÉS						
Diagnostique, travail de groupe & participation	I	I	I	I	I	N
Surveillance des risques et comportements	I	I	I	I	I	I
Planification et stratégie de communication	I	I	I	M	N	N
Évaluation participative	I	I	I	I	N	N
Ateliers "auto-organisés"	A					
TRAVAIL COMMUNAUTAIRE						
Cartographie des risques	I	I	I	I	N	N
Diagnostique des problèmes	I	I	I	I	N	N
Plan d'action	I	I	I	N	N	N
Actions exécutées	I	I	I	N	N	N
Stratégie de communication	I	N	N	N	N	N
Élaboration des matériaux de communication	I	N	N	N	N	N
Suivi et évaluation	M	M	M	M	M	N
Évaluation des points forts de la communauté						
ORGANIZATION & GESTION						
Présence des GTCs conduisant la stratégie	I	I	I	M	N	N
Personel *Aa*CP dans les GTC	I	I	I	M	N	N
Identification des ressources communautaires	M	I	N	N	N	N
Mobilisation des ressources extérieures	I	N	N	N	N	N
Lien avec l'école primaire établi						
Implication d'autres acteurs	A	A				

Table 2b. Composantes et sous-composantes de la stratégie mises en œuvre, modifiées, non mises en œuvre ou ajoutées par le Conseil Populaire 2 et les circonscriptions. Municipalité de La Lisa, La Havane, 2004-2007

COMPOSANTES DE LA STRATÉGIE	Conseil Populaire 3				
	12	13	14	15	16
SURVEILLANCE					
Identification des risques environnementaux	I	I	I	N	N
Identification des risques domiciliaires	I	I	I	I	I
Identification des comportements associés	I	I	N	N	N
RENFORCEMENT DES CAPACITÉS					
Diagnostique, travail de groupe & participation	M	N	N	N	N
Surveillance des risques et comportements	M	M	M	M	M
Planification et stratégie de communication	M	N	M	N	N
Évaluation participative	N	N	N	N	N
Ateliers "auto-organisés"					
TRAVAIL COMMUNAUTAIRE					
Cartographie des risques	I	I	I	N	N
Diagnostique des problèmes	I	I	I	N	N
Plan d'action	I	M	I	N	N
Actions exécutées	I	I	N	N	N
Stratégie de communication	N	N	N	N	N
Élaboration des matériaux de communication	N	N	N	N	N
Suivi et évaluation	N	N	N	N	N
Évaluation des points forts de la communauté					
ORGANIZATION & GESTION					
Présence des GTCs conduisant la stratégie	I	M	I	N	N
Personel *Aa*CP dans les GTC	M	M	M	N	N
Identification des ressources communautaires	N	N	N	N	N
Mobilisation des ressources extérieures	N	N	N	N	N
Lien avec l'école primaire établi					
Implication d'autres acteurs			A		

Table 2c. Composantes et sous-composantes de la stratégie mises en œuvre, modifiées, non mises en œuvre ou ajoutées par le Conseil Populaire 3 et les circonscriptions. Municipalité de La Lisa, La Havane, 2004-2007

À part la surveillance et le renforcement des capacités, toutes les composantes étaient significativement et fortement corrélées (coefficient de corrélation Spearman Ranketgt; 0,70, P etlt;0,01). Si une composante avait été mise en œuvre dans une circonscription, les autres composantes étaient également susceptibles de l'être. La corrélation entre la mise en œuvre des composantes du travail communautaire et de l'organisation et la gestion a atteint 0,95 tandis que la corrélation entre les composantes de la surveillance et du renforcement des capacités était plus faible (coefficient de 0,50).

Les sous-composantes très peu mises en œuvre ont été la mobilisation

des ressources internes et externes par les circonscriptions, la conception de stratégies de communication et l'élaboration de matériaux de communication locaux. Les principales modifications apportées ont porté sur la composition des groupes de travail communautaires, la modification de l'approche du renforcement des capacités, de la participation individualisée et des modifications dans la conception des activités de formation avec l'utilisation de techniques participatives plus adaptées aux caractéristiques des participant-e-s. Considérant l'ensemble de la stratégie, certaines activités ont été ajoutées, telles que l'incorporation à la stratégie des activités des écoles primaires, l'implication d'acteurs et actrices initialement non prévus, des ateliers de formation communautaire supplémentaires et l'évaluation des points forts de la communauté dans le cadre du diagnostic participatif initial.

Les difficultés rencontrées lors de la mise en œuvre étaient liées à la formation et à la mobilisation de compétences appropriées, au temps disponible, au manque de soutien et d'engagement du gouvernement local et des autorités sanitaires, au manque de motivation des leaders locaux, d'intégration des acteurs et actrices et de mobilisation de ressources. L'étude a montré une grande variabilité de la fidélité dans la mise en œuvre. Cette variabilité s'explique, en grande partie, par la complexité de la stratégie, qui a parfois perturbé les pratiques courantes de contrôle des vecteurs et le manque de connaissances des responsables de la mise en œuvre en ce qui concernait les principes de base de la stratégie. Cependant, la variation de la fidélité dans la mise en œuvre n'a pas entraîné une perte substantielle d'efficacité (Castro *et al.* 2012).

Analyse réflexive

La typologie de Rebchook (*et al.* 2006) s'est révélée appropriée pour l'évaluation de la fidélité et de l'adaptation, mais elle a montré des limites. Elle ne nous a pas permis de savoir quelles modifications introduites dans la stratégie ont contribué ou non à l'atteinte des résultats. L'affirmation de Rebchook (*et al.* 2006) selon laquelle l'ajout de nouvelles composantes ou sous-composantes à un modèle initial ne menace pas *a priori* la fidélité est également discutable. En effet, cet ajout pourrait être en contradiction avec la théorie du changement sous-jacent à l'intervention. En outre, la typologie

ne nous a pas procuré une vue très détaillée de la fidélité de l'intervention concernant le contenu, la dose et la couverture. Par conséquent, nous avons confronté notre recherche empirique avec la littérature existante sur la fidélité. Nous avons identifié le cadre d'analyse de la fidélité de la mise en œuvre de Carroll et ses collègues (2007) comme celui qui pouvait potentiellement répondre à notre préoccupation d'évaluer l'adaptation dans le contexte de la fidélité (Pérez *et al.* 2016). Mais, il nous fallait l'adapter pour tenir compte des interventions adaptables de santé mondiale. Ainsi, nous proposons quelques modifications à ce cadre avec des implications pratiques pour l'évaluation de la fidélité et de l'adaptation (Pérez *et al.* 2016).

Notre cadre modifié pour l'analyse de la fidélité de la mise en œuvre retient les idées de Carroll (*et al.* 2007) en ce qui concerne l'évaluation des effets atteints et l'analyse des composantes pour identifier celles essentielles à une intervention. En effet, l'évaluation de la mise en œuvre et l'atteinte d'un équilibre adéquat entre fidélité et adaptation n'ont de sens que dans le contexte de l'appréciation de ses effets (Backer 2002). Cependant, nous avons amélioré la représentation graphique du rôle de la fidélité et de l'évaluation des effets dans l'identification des composantes essentielles d'une intervention (Figure 2).

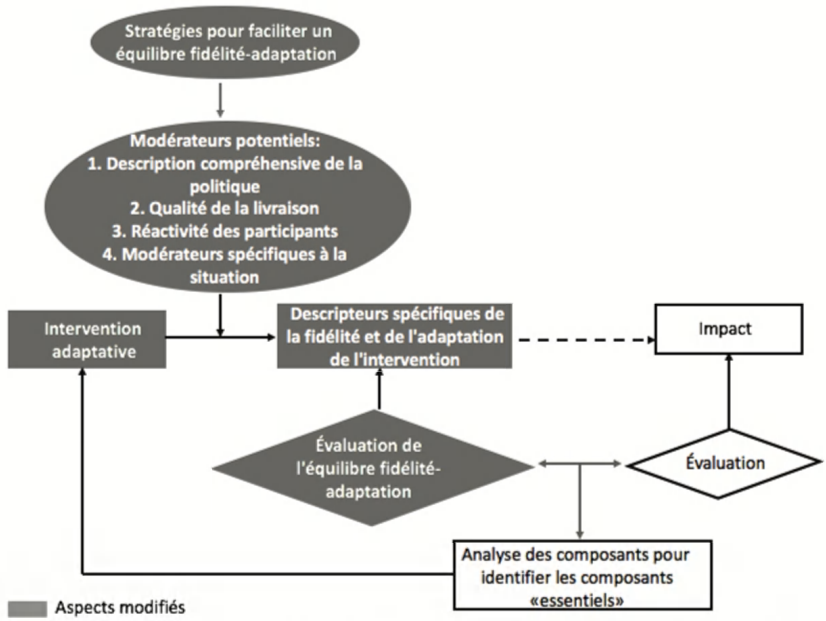


Figure 2. Cadre conceptuel modifié de Carroll et al. 2007 pour l'évaluation de la mise en œuvre

Carroll et ses collègues (2007) ont établi l'adhérence comme la seule mesure de base de la fidélité de la mise en œuvre. Comme nous l'avons appris de notre recherche empirique sur la fidélité (Pérez *et al.* 2010), la nature des adaptations doit être consciemment prise en compte pour apprécier leur effet sur l'efficacité. De plus, les sous-catégories de l'adhérence peuvent ne pas être toujours suffisantes pour évaluer chaque intervention. D'autres aspects pourraient devoir être pris en compte. Ainsi, nous suggérons d'évaluer les descripteurs spécifiques de l'intervention aussi bien pour la fidélité que pour l'adaptation, au lieu de considérer seulement les sous-catégories fixes de l'adhérence.

Pour avancer dans cette direction, les concepteurs et conceptrices d'intervention ont besoin 1) d'avoir une idée claire du (des) résultat (s) attendu (s), 2) de rendre explicites les principes de fonctionnement ou la théorie du changement de l'intervention, 3) d'être en mesure de les présenter sous la forme de descripteurs spécifiques de la fidélité et 4) d'établir des questions

pour identifier les adaptations en rapport avec la description de l'intervention. Les descripteurs spécifiques des adaptations de l'intervention sont développés en répondant à ces questions. Une étape supplémentaire consiste à déterminer dans quelle mesure les adaptations identifiées affectent les principes de fonctionnement d'une composante et/ou de l'intervention dans son ensemble.

Un exemple de ce raisonnement peut être proposé en utilisant la composante du renforcement des capacités de notre stratégie. Le résultat attendu de cette composante était que les participant-e-s acquièrent les connaissances et compétences nécessaires pour modifier les relations de pouvoir existantes entre eux. Les principes de fonctionnement ont été établis en utilisant le modèle pédagogique de l'éducation populaire (Freire 1972). Les participant-e-s à la formation doivent acquérir des connaissances et des compétences spécifiques grâce à un processus d'apprentissage participatif (groupe d'apprentissage). Le groupe d'apprentissage doit impliquer des intervenant-e-s qui doivent changer leurs relations de pouvoir. Les descripteurs spécifiques de la fidélité pour le renforcement des capacités et les questions pour identifier les éventuelles adaptations sont dans la Table 3.

Table 3. Exemple de descripteurs spécifiques de la fidélité et de questions pour identifier les adaptations pour la composante renforcement des capacités de la stratégie de renforcement de la capacité à agir pour la lutte contre le vecteur de la dengue. Municipalité de La Lisa, La Havane, 2004-2007.

Descripteurs spécifiques de la fidélité	**Questions pour identifier des adaptations**
Quoi : développement de connaissances et de compétences sur quatre sujets : 1) diagnostic, travail en groupe et participation; 2) surveillance des risques et des comportements; 3) plans d'action et stratégie de communication; et 4) évaluation participative	Le contenu de la formation a-t-il été changé? Comment? Un sujet a-t-il été supprimé? Lequel? Pourquoi? Un sujet a-t-il été remplacé? Par quoi ? Pourquoi? Un sujet a-t-il été ajouté? Lequel? Pourquoi?
Comment : à travers un atelier basé sur les principes du modèle pédagogique de l'éducation populaire : l'objectif est que les participant-e-s réfléchissent et, par conséquent, transforment leur réalité en utilisant une logique dialectique entre la théorie et la pratique et les méthodes d'apprentissage participatives et bases sur l'expérience.	Est-ce qu'un principe du modèle pédagogique a été adapté : objectifs, logique, méthodes d'apprentissage? Lequel? Comment? Pourquoi? Le modèle pédagogique a-t-il été remplacé par un autre? Par lequel Pourquoi?
À quelle fréquence : un atelier de quatre heures pour chaque sujet étalé sur une période de trois mois	Une adaptation a-t-elle été introduite dans la fréquence de la formation : nombre de sessions, nombre d'heures par session ? Comment? Pourquoi? Les séances de l'atelier se sont-elles réduites au fil du temps? Comment? Pourquoi? Y a-t-il eu une adaptation introduite dans la durée et la période prévue pour la formation? Comment? Pourquoi?
À qui : un groupe d'apprentissage composé de trois à cinq acteurs et actrices appartenant à au moins trois communautés et entretenant différentes relations de pouvoir entre eux, en rapport avec les activités de lutte contre le vecteur de la dengue.	Le groupe d'apprentissage a-t-il été adapté : la quantité de participant-e-s, le rôle des parties prenantes en relation avec les activités de lutte contre le vecteur de la dengue? Comment? Pourquoi? Le groupe d'apprentissage a-t-il été remplacé par une autre stratégie d'enseignement? Par laquelle ? Pourquoi?
Par qui : Les facilitateurs et facilitatrices précédemment formés sur la base des principes du modèle pédagogique de l'éducation populaire.	Un facilitateur n'a-t-il pas été formé? Pourquoi? Est-ce qu'un principe du modèle pédagogique a été adapté lors de la formation des facilitateurs? Lequel? Comment? Pourquoi? Le modèle pédagogique a-t-il été remplacé par un autre? Par lequel ? Pourquoi?

Spécifications liées au contexte : Trois conseils populaires (CP) sont impliqués dans le projet. Il existe cinq à six circonscriptions sélectionnées au hasard par CP. Un soutien matériel pour la formation est fourni : par exemple, des directives écrites sur la façon de mener un atelier d'éducation populaire, des conseils méthodologiques aux facilitateurs par au moins un concepteur de la stratégie.	Y a-t-il eu un changement dans le nombre de CP? Pourquoi? Y a-t-il eu un changement dans le nombre de circonscriptions impliquées? Une circonscription a-t-elle été remplacée? Comment? Pourquoi? Des modifications ont-elles été apportées au soutien méthodologique : la mise à disposition de lignes directrices, le contenu des lignes directrices, le conseil méthodologique? Comment? Pourquoi?

Remarquons que les descripteurs spécifiques de la fidélité fournissent une description complète de l'intervention comme prévu, avec des détails sur le contenu, les processus (« quoi », « comment », « à quelle fréquence », « à qui » et « par qui ») et les spécifications liées au contexte de la mise en œuvre. Supposons que, lors de l'application des questions aux données empiriques, les adaptations identifiées concernant le renforcement des capacités avaient été modifiées de trois manières différentes : 1) le sujet « diagnostic, travail en groupe et participation » a été supprimé, 2) un atelier sur les stratégies de communication a été ajouté et 3) les méthodes d'apprentissage ont été adaptées aux caractéristiques des participant-e-s. En prenant en compte nos principes de fonctionnement, nous devons nous demander laquelle des trois adaptations pourrait influencer les effets escomptés. De notre point de vue, il n'est pas possible de fournir aux participant-e-s la capacité de participer, sans leur enseigner le sujet supprimé. À l'inverse, nous ne considérons pas que l'une des deux autres adaptations menace les principes de fonctionnement du renforcement des capacités et ses résultats.

Grâce à cette analyse des adaptations qui surviennent dans la conduite d'une intervention, des pistes de solutions pour une amélioration future de la mise en œuvre peuvent être identifiées. Cela permet d'aider les concepteurs d'interventions à identifier les adaptations non prédéfinies, qui pourraient améliorer l'élaboration de l'intervention et donc son efficacité. Une fois qu'une adaptation positive est identifiée, l'intervention pourrait passer par un nouveau cycle de conception, de mise en œuvre et d'analyse. Cela nécessiterait des mécanismes de rétroaction.

La nécessité d'identifier davantage les sources potentielles de variabilité dans la mise en œuvre a été soulignée par certains auteurs (Arai *et al.* 2005; Roen *et al.* 2006; Belaid et Ridde 2015). Carroll et ses collègues (2007)

reconnaissent que le niveau de fidélité atteint est influencé par des facteurs de modération potentiels qui ne sont pas nécessairement indépendants. Hasson et ses collègues (2012) ont d'autre part souligné l'importance d'autres mécanismes et facteurs influençant la fidélité de la mise en œuvre quand ils ont expérimenté le cadre de Carroll et ses collègues (2007). Notre cadre modifié maintient l'exhaustivité de la description des politiques, la qualité de l'exécution et la réactivité des participant-e-s en tant que modératrices et modérateurs potentiels de base. Cependant, il permet d'en inclure d'autres spécifiques au contexte ou à la situation.

En outre, nous sommes d'accord avec Carroll et ses collègues (2007) sur le fait que les stratégies de facilitation pourraient influencer les modératrices et les modérateurs potentiels du niveau de fidélité atteint. Par exemple, la fourniture de manuels et de formation aux responsables de la mise en œuvre pourrait améliorer la qualité de l'exécution d'une intervention. Cependant, dans notre cadre modifié, ces stratégies sont mises en place non pas dans le but d'atteindre une adhérence stricte, mais pour contribuer à un équilibre adéquat entre adaptation et fidélité. Une fois que les adaptations ont été identifiées, les stratégies de facilitation s'adresseront uniquement à celles jugées inadéquates. Dans le cas d'interventions adaptables, cela permet également d'avoir une vue globale sur les adaptations survenues.

Dans la pratique, l'objectif d'atteindre un équilibre fidélité-adaptation implique leur interdépendance. La nécessité d'un tel équilibre, souligné par Backer (2002) et von Thiele Schwarz (*et al.* 2014), est fortement appuyée par les résultats de notre recherche empirique sur la fidélité (Pérez *et al.* 2010) et nos recherches complémentaires sur la diffusion de la stratégie de renforcement des capacités à agir pour la prévention de la dengue (Pérez *et al.* 2013).

Références clés

Rebchook, M., Kegeles, S., Huebner, D. et TRIP Research Team. (2006). Translating research into practice: the dissemination and initial implementation of an evidence-based HIV Prevention Program. *AIDS Education and Prevention*, 18(Suppl. A), 119-136.

Cet article présente les résultats d'une recherche empirique sur la mise à l'échelle d'un programme de prévention du VIH. Un de ses objectifs est

d'explorer comment les parties prenantes essaient de mettre en œuvre le programme. Le papier analyse les types de réinventions/adaptations apportées et leur implication sur la fidélité.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J. et Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2, 40-48.

Cet article fournit le cadre conceptuel le plus compréhensif pour l'évaluation de la fidélité dans la mise en œuvre. Ce cadre s'est révélé par la suite être un outil d'évaluation utile pour l'évaluation de la fidélité d'interventions complexes.

Rogers, E. (2003). *Diffusion of innovations* (5e éd.). New York: New York Free Press.

La cinquième édition du livre où l'auteur explique comment de nouvelles idées et technologies se diffusent parmi les membres d'un système social et sont mis en pratique. L'ouvrage est basé sur l'observation de régularités dans la diffusion de différentes innovations parmi différentes cultures et usagers et contient des développements théoriques élaborés sur des décades.

Références

Arai, L., Roen, K., Roberts, H. et Popay, J. (2005). It might work in Oklahoma but will it work in Sout? Oakhampton: Context and implementation in the effectiveness literature on domestic smoke detectors. *Injury Prevention*, 11(3),148-151.

Backer, T. (2002). *Finding the balance: Program fidelity and adaptation in substance abuse prevention: A state-of-the-art review*. Rockville, M. (ed.). Center for Substance Abuse Prevention.

Belaid, L. et Ridde, V. (2015). Contextual factors as a key to understanding the heterogeneity of effects of a maternal health policy in Burkina Faso?. *Health Policy and Planning*, 30(3), 309-321.

Bellg, A., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D., Ory, M.,Ogedegbe, G., Orwig, D., Ernst, D. et Czajkowski, S. (2004). Enhancing treatment fidelity in health behaviour change studies: Best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology*, 23(5), 443-451.

Borrelli, B. (2011). The assessment, monitoring, and enhancement of treatment fidelity in public health clinical trials. *Journal of Public Health Dentistry*, 71, 52-63.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J. et Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2, 40-48. http://www.implementationscience.com/content/2/1/40

Castro, M., Pérez, D., Pérez, K., Polo, V., Lopez, M. et Sanchez, L. (2008). Contextualización de una estrategia comunitaria integrada para la prevención del dengue. *Revista Cubana de Medicina Tropical*, 60, 83-91.

Castro, M., Sanchez, L., Pérez, D., Carbonell, N., Lefèvre, P., Vanlerberghe, V. et Van der Stuyft, P. (2012). A community empowerment strategy embedded in a routine dengue vector control programme: a cluster randomised controlled trial. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 106(5), 315-321.

Collins, L., Murphy, S. et Bierman, K. (2004). A conceptual framework for adaptive preventive interventions. *Prevention Science*, 5(3), 185-196.

Creswell, J. et Clark, V. P. (2011). *Design and conducting mixed methods research*. Washington, DC: Sage Publications.

Dane, A. et Schneider, B. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control?. *Clinical Psychology Review*, 18(1), 23-45.

Dusenbury, L., Brannigan, R., Falco, M. et Hansen, W. (2003). A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Education Research*, 18(2), 237-256.

Fixsen, D., Naoom, S., Blase, K., Friedman, R. et Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa, FL: University of South Florida.

Freire, P. (1972). *Pedagogy of the Oppressed*. Londres: Sheed and Ward.

Gagliardi, A., Straus, S., Shojania, K. et Urbach, D. (2014). Multiple interacting factors influence adherence and outcomes associated with surgical safety checklists: a qualitative study. *PLoS One*, 9(9): e108585. https://doi.org/10.1371/journal.pone.0108585

Guzman, M., Rodriguez-Roche, R. et Kouri, G. (2006). Dengue and dengue hemorrhagic fever. *Giornale Italiano Di Medicina Tropicale*, 11, 1-2.

Hasson, H., Blomberg, S. et Dunér, A. (2012). Fidelity and moderating factors in complex interventions: a case study of a continuum of care program for frail elderly people in health and social care. *Implementation Science*,7, 23-33.
https://doi.org/10.1186/1748-5908-7-23

Hernandez, M. et Hodges, S. (2003). Building upon the theory of change for systems of care. *Journal of Emotional and Behavioral Disorders*, 11(1), 19-26.

Kay, B. et Nam, V. (2005). New strategy against Aedes aegypti in Vietnam. *The Lancet*, 365(9459), 613-617.

Kay, H., Thanh, T. T., Le, N., Quy, T., Nam, V., Hang, P., ... et Ryan, P. (2010). Sustainability and cost of a community-based strategy against Aedes aegypti in northern and central Vietnam. *American Journal of Tropical Medicine and Hygiene*, 82(5), 822-830.

Kouri, G., Guzman, M., Bravo, J. et Triana, C. (1989). Dengue haemorrhagic fever/dengue shock syndrome : lessons from the Cuban epidemic, 1981. *Bulletin of World Health Organization*, 67(4), 375-380.

Parks, W. et Lloyd, L. (2004). *Planning social mobilization and communication for dengue fever prevention and control: A step-by-step guide*. Genève: World Health Organization.

Pérez, D., Lefèvre, P., Castro, M., Sanchez, L., Toledo, M., Vanlerberghe, V. et Van der Stuyft, P. (2010). Process-oriented fidelity research assists in evaluation, adjustment and scaling up of community-based interventions. *Health Policy and Planning*, 26(5), 413-422.

Pérez, D., Lefèvre, P., Castro, M., Toledo, M., Zamora, G., Bonet, M. et Van der Stuyft, P. (2013). Diffusion of community empowerment strategies for Aedes aegypti control in Cuba: A muddling through experience. *Social Science and Medicine*, 84, 44-52.

Pérez, D., Van der Stuyft, P., Zabala, M. et Lefèvre, P. (2016). A modified theoretical framework to assess implementation fidelity of adaptive public health interventions. *Implementation Science*,11, 106.
https://doi.org/10.1186/s13012-016-0457-8

Peters, D., Tran, N. et Adam, T. (2013). *Implementation Research in Health: A Practical Guide*. Genève: World Health Organization.

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R. et Hensley, M. (2011). Outcomes for implementation research: conceptual distinctions, measurement, challenges and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research*, 38(2), 65-76. https://doi.org/10.1007/s10488-010-0319-7

Rebchook, M., Kegeles, S., Huebner, D. et TRIP Research Team. (2006). Translating research into practice: the dissemination and initial implementation of an evidence-based HIV Prevention Program. *AIDS Education and Prevention*, 18(Suppl. A), 119-136.

Rifkin, S. (1996). Paradigms lost: Toward a new understanding of community participation in health programmes. *Acta Tropica*, 61(2), 79-92.

Rifkin, S. (2014). Examining the links between community participation and health outcomes: a review of the literature. *Health Policy and Planning*, 29(Suppl. 2), 98-106.

Roen, K., Arai, L., Roberts, H. et Popay, J. (2006). Extending systematic reviews to include evidence on implementation: methodological work on a review of community-based initiatives to prevent injuries. *Social Science and Medicine*, 63(4), 1060-1071.

Rogers, E. (2003). *Diffusion of innovations* (5e éd.). New York: New York Free Press.

Sanchez, L., Pérez, D., Cruz, G., Castro, M., Kouri, G., Shkedy, Z., Vanlerberghe, V. et Van der Stuyft, P. (2009). Intersectoral coordination, community empowerment and dengue prevention: six years of controlled interventions in Playa Municipality, Havana, Cuba. *Tropical Medicine and International Health*,14(11), 1356-1364.

Schwarz, U. V., Hasson, H. et Lindfors, P. (2014). Applying a fidelity framework to understand adaptations in an occupational health intervention. *Work*, 51(2), 195-203.

Stoddard, S., Forshey, B., Morrison, A., Paz-Soldan, V., Vazquez-Prokopec, G., Astete, H., Reiner, R.C., Vilcarromero, S., Elder, J. P., Halsey, E. S., Kochel, T. J., Kitron, U. et Scott, T. (2013). House-to-house human movement drives

dengue virus transmission. *Proceedings from the National Academy of Sciences of the United States of America*, 110(3), 994-999.

Sundell, K., Beelmann, A. H et Schwarz, U. V. (2015). Novel programs, international adoptions, or contextual adaptations?: Meta-analytical results from German and Swedish intervention research. *Journal of Clinical Child and Adolescent Psychology*, 45:6, 784-796. http://doi:10.1080/15374416.2015.1020540

Toledo, M., Baly, A., Vanlerberghe, V., Rodriguez, M., Benitez, J., Duvergel, J. et Van der Stuyft, P. (2008). The unbearable lightness of technocratic efforts at dengue control. *Tropical Medicine and International Health*, 13(5), 728-736.

Toledo, M., Vanlerberghe, V., Baly, A., Ceballos, E., Valdes, L., Searret, M., Boelaert, M. et Van der Stuyft, P. (2007). Towards active community participation in dengue vector control: results from action research in Santiago de Cuba, Cuba. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 101, 56-63.

Vanlerberghe, V., Toledo, M., Rodriguez, M., Gomez, D., Baly, A., Benitez, J. et Van der Stuyft, P. (2009). Community involvement in dengue vector control: cluster randomised trial. *British Medical Journal*, 338 :b1959. https://www.bmj.com/content/338/bmj.b1959

Résumé / Resumen / Abstract

Ce chapitre porte sur l'évaluation de la fidélité dans la mise en œuvre et l'adaptation des interventions de santé mondiales. La fidélité, ou le degré avec lequel une intervention est mise en œuvre comme prévu par ses concepteurs. Elle est particulièrement destinée à assurer que l'intervention maintienne ses effets prévus. L'adaptation, au contraire, est la modification du modèle d'origine d'une intervention par les acteurs et actrices de sa mise en œuvre. Cinq dimensions ont été mises en avant pour mesurer la fidélité : l'adhérence, la dose, la qualité de la prestation, la réactivité des participant-e-s et la différentiation du programme. Cependant, peu de recherches ou de conseils pratiques portent sur la façon d'adapter une intervention, tout en maintenant les mécanismes essentiels à son efficacité. Les interventions peuvent être un mélange de fidélité et d'adaptation et l'évaluation doit

examiner les deux aspects. Nous avons confronté notre recherche empirique avec la littérature existante sur la fidélité et en avons déduit que le cadre conceptuel proposé par Carroll permettait d'évaluer l'adaptation dans le contexte de la fidélité. Nous proposons des modifications à ce cadre pour accommoder les interventions en santé mondiales adaptables ainsi que des implications pratiques.

***

Este capítulo se centra en la evaluación de la fidelidad en la implementación y la adaptación de intervenciones en salud global. La fidelidad, o el grado en que una intervención se implementa según lo previsto por sus diseñadores, está especialmente destinada a garantizar que la intervención mantenga los efectos previstos. La adaptación, por el contrario, es la modificación del modelo original de una intervención por parte de los actores involucrados en su implementación. Existen cinco dimensiones para medir la fidelidad: adherencia, dosis, calidad del servicio, respuesta de los participantes y diferenciación del programa. Sin embargo, existe poca investigación o consejos prácticos sobre cómo adaptar una intervención manteniendo los mecanismos que garantizan su eficacia. La implementación de intervenciones puede ser una mezcla de fidelidad y adaptación, por lo que en la evaluación deben considerarse ambos aspectos. Comparamos nuestra investigación empírica con la literatura existente sobre fidelidad. Como resultado consideramos que el marco conceptual de Carroll y colaboradores permitiría evaluar adaptación en el contexto de fidelidad. Por ello proponemos modificaciones a este marco teórico para adecuarlo a las intervenciones adaptativas en salud global con implicaciones prácticas para la evaluación de fidelidad y adaptación.

***

This chapter is about the evaluation of the implementation fidelity and adaptation while building evidence for global health interventions. This issue is particularly relevant for adaptive interventions. Fidelity or the degree to which an intervention is implemented as intended by its developers is an implementation outcome that is particularly meant to ensure that the intervention maintains its intended effects. Adaptation, on the opposite is the process of bringing changes to the original design of an intervention by its implementers or users. Five dimensions have been forwarded to measure fidelity: adherence, dose, quality of delivery, participant responsiveness, and

program differentiation. However, there has been little research or practical advice on how to adapt an intervention to maintain its effective ingredients and mechanisms. Interventions can be a blend of both fidelity and adaptation; thus, evaluation must carefully examine both aspects. We confronted our empirical research with existing literature on fidelity. As a result, we considered that the framework for implementation fidelity proposed by Carroll and colleagues responds to our concern to assess adaptation in the context of fidelity. We proposed modifications to the framework to accommodate adaptable global health interventions with practical implications for the evaluation of fidelity and adaptation.

***

Dennis Pérez est une sociologue cubaine avec une spécialisation en éducation populaire et en recherche-action participative. Elle détient un Master en Contrôle des Maladies et en Développement social, un Doctorat en Sociologie (Université de la Havana) et en Sciences de la Santé (Université de Gand). Depuis 1997, elle travaille comme chercheuse dans l'Unité d'Épidémiologie de l'Institut de Médecine Tropicale Pedro Kouri à La Havane. Actuellement elle dispose d'une bourse postdoctorale à l'Institut de Recherche en Santé publique de l'Université de Montréal. Ses principales recherches portent sur l'évaluation des processus de mise en œuvre ; la fidélité dans la réplication et la diffusion de stratégies pour le contrôle de la dengue basées sur le renforcement des capacités à agir ; les développements théoriques et méthodologiques afférents. Elle s'intéresse également aux études qualitatives portant sur les perceptions et l'acceptabilité par les populations d'interventions et des services de santé.

Marta Castro est médecin avec une spécialisation en psychiatrie pédiatrique. Elle détient également un Master et un Doctorat en Épidémiologie. Elle travaille actuellement comme chercheuse dans l'Unité d'Épidémiologie de l'Institut de Médecine Tropicale Pedro Kourià La Havane essentiellement sur des projets d'épidémiologie sociale en relation avec la dengue, le V.I.H.et la TB. Ses projets récents portent sur le design, la mise en œuvre et l'évaluation de programmes communautaires de prévention et de contrôle de la dengue, avec une attention spéciale prêtée à comprendre les déterminants sociaux de la santé et la promotion de la participation. Elle a également conduit une recherche qualitative sur le V.I.H.parmi les enfants et adolescents à Cuba.

Pierre Lefèvre est Docteur en Sociologie, diplômé de l'Université Libre de Bruxelles. Entre 1989 et 2017, il a travaillé comme chercheur au Département de Santé publique de l'Institut de Médecine Tropicale d'Anvers, Belgique. D'abord à l'Unité de nutrition et ensuite à l'Unité d'épidémiologie et de contrôle des maladies. Il est actuellement consultant indépendant en santé publique internationale. Ses recherches ont porté sur l'évaluation pluraliste (thèse de Doctorat), la planification participative, la participation communautaire, la transition nutritionnelle, la croissance et le développement de l'enfant, et les perceptions de diverses maladies et interventions par les populations dans de nombreux pays. Il a également appuyé méthodologiquement (design, analyse) les composantes qualitatives de nombreux projets de recherche épidémiologiques. Ses centres d'intérêt actuels sont l'institutionnalisation de la participation communautaire dans les programmes de contrôle des vecteurs, les études d'acceptabilité et, plus généralement, la science de la mise en œuvre.

Citation

Dennis Pérez, Marta Castro et Pierre Lefèvre. (2019). L'évaluation de la pérennité. Une intervention de financement basé sur les résultats au Mali. In *Évaluation des interventions de santé mondiale. Méthodes avancées*. Sous la direction de Valéry Ridde et Christian Dagenais, pp. 403-432. Québec : Éditions science et bien commun et Marseille : IRD Éditions.

17. L'évaluation réaliste

L'exemple de l'adoption d'une politique publique de santé au Bénin

JEAN-PAUL DOSSOU ET BRUNO MARCHAL

L'évaluation réaliste

Définition de la méthode

Cette approche d'évaluation fondée sur la théorie cherche à expliciter de manière systématique les mécanismes causaux des résultats d'une intervention pour différentes populations et dans différents contextes. Elle permet de répondre non seulement à la question classique « Est-ce qu'une intervention fonctionne ou non pour entrainer tel effet? », mais surtout à des questions importantes pour la prise de décision comme « Pourquoi une intervention fonctionne ou non? Comment? Pour qui et dans quel contexte? ».US

Forces de la méthode

- C'est une approche émergente; les personnes qui maîtrisent sont de plus en plus demandées.
- Elle permet de répondre à des questions d'évaluation impossible à aborder avec d'autres approches comme : Comment, pour qui et dans quel contexte, une intervention marche ou non, et pourquoi?
- Elle met les actrices et acteurs opérationnels au cœur de l'évaluation. Ce focus valorisant et capacitant fait du processus d'évaluation un outil au service d'une meilleure redistribution des pouvoirs.
- Elle a le potentiel de stimuler et de renforcer la culture locale de l'apprentissage en favorisant une explicitation systématique et une confrontation des hypothèses habituellement implicites qui sous-tendent les interventions et les pratiques.

Défis de la méthode

- Elle dépend de la richesse des connaissances de l'évaluateur ou de l'évaluatrice sur le programme en évaluation, les acteurs et actrices et les contextes de mise en œuvre ; elle nécessite, de fait, des ressources relativement importantes.
- Elle nécessite un effort de synthèse important et de la créativité pour communiquer efficacement les résultats.
- Le critère de qualité des résultats des évaluations réalistes est la plausibilité. Les commanditaires d'évaluation soucieux d'attribuer les résultats à un programme donné verraient ceci potentiellement comme une limite.
- Elle reste une approche en construction avec un débat méthodologique encore vif. Il peut être parfois difficile de trouver des repères méthodologiques clairs, précis et établis pour l'étudiant-e réaliste débutant-e.

Présentation de la méthode, utilité et défis d'application

L'évaluation réaliste est une approche d'évaluation fondée sur la théorie (Stern *et al.* 2012) Elle permet de répondre non seulement à la question classique « Est-ce qu'une intervention I fonctionne ou non pour entrainer un effet E? », mais surtout à des questions importantes pour la prise de décision comme « Pourquoi une intervention I fonctionne-t-elle ou non? Comment? Pour qui et dans quel contexte? » (Ridde et Dagenais 2012). Elle cherche à expliciter de manière systématique les mécanismes causaux des résultats d'une intervention pour différentes populations et dans différents contextes et elle permet ainsi l'adaptation optimale des programmes sociaux à différents groupes de personnes. Elle constitue de fait, une excellente approche pour promouvoir l'équité.

L'évaluation réaliste est plus adaptée aux interventions complexes mises en œuvre dans des systèmes ouverts en mutation permanente (Marchal *et al.* 2012). Elle convient à l'évaluation des nouvelles initiatives, des interventions pilotes et des interventions aux résultats mitigés, dont on ignore les mécanismes de succès ou d'échec afin de comprendre les différences (Westhorp 2014). L'évaluation réaliste convient aussi à l'évaluation d'interventions dont l'approche fait l'objet d'un faible consensus entre les parties prenantes, sur laquelle il existe très peu de connaissances à propos des options possibles ou dans laquelle l'expertise est limitée. Dans ce cas, l'évaluation réaliste permet d'adopter une démarche d'apprentissage systématique sur des questions clés importantes pour guider la politique dans son processus de développement, de mise en œuvre et d'amélioration progressive.

Cette forme d'évaluation rompt fondamentalement avec les approches classiques d'évaluation parce qu'elle s'appuie sur les concepts de la causalité générative et du courant philosophique du réalisme (Pawson et Tilley 1997). Selon le modèle de causalité générative, une action est causale si, et seulement si, ses effets sont déclenchés par un mécanisme qui agit dans un contexte spécifique. Ce mécanisme causal est l'objet central des travaux de l'évaluateur ou de l'évaluatrice réaliste[1]. L'intervention n'est pas perçue

1. Malgré les nuances entre le terme anglais « *realist* » et le terme français « réaliste », nous utilisons dans ce chapitre le mot « réaliste » comme une traduction de « *realist* ». Nous nous appuyons sur l'utilisation déjà répandue dans le langage scientifique

comme causale : elle apporte des ressources qui, en interaction avec les ressources préexistantes dans le contexte, peuvent générer ou non un mécanisme causal déclenchant un effet. La causalité générative se distingue ainsi du modèle de causalité successionniste qui stipule qu'un effet E peut être attribué à une intervention I si on peut démontrer la conjonction constante entre E et I. La nuance entre ces deux modèles de causalité est essentielle pour définir et comprendre l'évaluation réaliste (Van Belle *et al.* 2016).

L'évaluation réaliste s'inscrit dans le courant philosophique du réalisme, qui reconnait que la réalité existe indépendamment des interprétations ou de la connaissance que l'humain a de cette réalité (Audi 1999). Le réalisme philosophique est distinct du positivisme et du constructivisme qui sont à la base des approches d'évaluation des programmes les plus utilisées (Pawson 2006; Westhorp *et al.* 2011) Pour mieux comprendre l'évaluation réaliste, il est essentiel de connaitre ces courants et leurs implications en matière d'évaluation. Un premier livre de Ridde et Dagenais discute longuement de ces fondements théoriques et de leurs implications pratiques (Ridde et Dagenais 2012).

Dans ce chapitre, nous présentons l'application de cette approche à l'évaluation de l'adoption d'une politique publique de santé (Goodnough *et al.* 1993). L'adoption s'inscrit dans la mise en œuvre, une phase complexe du cycle des politiques publiques qui est l'une des plus cruciales, mais malheureusement l'une des moins étudiées en Afrique subsaharienne (Ridde et Olivier de Sardan 2017). Nous présentons la méthodologie de cet exercice en suivant le cycle de l'évaluation réaliste (Marchal *et al.* 2012). Nous présentons les étapes du cycle et l'application que nous en avons faite. Une partie de nos résultats sera présentée pour illustrer la manière dont elles ont été « construites » à travers les différentes étapes du cycle. Finalement, nous discutons quelques leçons apprises.

francophone de l'expression « évaluation réaliste » comme une traduction de « *realist evaluation* ».

Contexte et programme évalué : la politique de gratuité de la césarienne au Bénin

Le Bénin a introduit en avril 2009 une politique d'exemption de paiement des frais de soins à l'attention des femmes nécessitant une césarienne dans 48 hôpitaux répartis sur tout le pays. Ces hôpitaux doivent offrir la césarienne sans perception de frais directs auprès des patientes. Le gouvernement rembourse en retour aux hôpitaux un montant forfaitaire de 100 000 FCFA (178 $US) par césarienne réalisée. Les frais de référence avant la césarienne, la consultation à l'admission dans l'hôpital, l'acte de césarienne, les médicaments et les consommables, l'hospitalisation et les soins postopératoires sont théoriquement pris en compte selon le décret qui institue la politique (Centre de Recherche en Reproduction humaine et en Démographie 2013). Une évaluation préliminaire avait mesuré les résultats de la mise en œuvre dans sept hôpitaux et a montré que le nombre d'éléments du paquet de gratuité, réellement offert de manière entièrement gratuite par les hôpitaux, varie de 1 sur 8 (Hôpital 2) à 5 sur 8 (Hôpital 5) (tableau 1). Cette variabilité dans l'adoption des politiques est un défi classique observé dans la mise en œuvre de plusieurs politiques de santé en Afrique subsaharienne.

Tableau 1. Adoption variable de la politique de gratuité de la césarienne dans 7 hôpitaux au Bénin, 2012

	Hôp. 1	Hôp. 2 (Cas 1)	Hôp. 3	Hôp. 4	Hôp. 5 (Cas 2)	Hôp. 6	Hôp. 7	Nombre d'hôpitaux offrant l'item gratuitement
Transfert dans le district sanitaire								0/7
Abord veineux avec soluté avant le transfert								0/7
Consultation								2/7
Acte chirurgical								6/7
Médicaments								1/7
Consommables								2/7
Hospitalisation								5/7
Contrôle postopératoire								6/7
Items gratuits sur nombre total d'items	3/8	1/8	2/8	3/8	5/8	4/8	4/8	

Source : Centre de Recherche en Reproduction humaine et en Démographie 2014

- Case de couleur noire : Totalement payant
- Case de couleur grise : Partiellement gratuit
- Case de couleur blanche : Totalement gratuit

Protocole d'évaluation : le cycle de l'évaluation réaliste

L'évaluation réaliste adopte une démarche cyclique qui commence avec une théorie de programme initiale et finit avec une théorie de programme raffinée. Ces théories sont des théories de moyenne portée (TMP), définies selon Merton (1968, p. 39) comme des théories

> dont le niveau se situe entre les hypothèses de travail, mineures, mais nécessaires, qui se développent au quotidien dans le cadre de la recherche, et les efforts systématiques d'intégration pour développer une théorie unitaire qui expliquerait toutes les uniformités observées du comportement social, de l'organisation sociale et du changement social ».

Le processus réaliste vise ainsi à produire des théories de programme de type TMP de plus en plus raffinées et de mieux en mieux aptes à être appliquées à des contextes de plus en plus variés et plus grands, ou encore à expliquer de manière plus fine et plus précise un phénomène social donné (Marchal *et al.* 2012).

Étape 1 du cycle de l'évaluation réaliste : formuler la question de recherche

La démarche pour formuler la question d'évaluation est similaire à celle de toutes les autres formes d'évaluation, à la différence que la perspective de l'évaluation réaliste élargit considérablement l'éventail des questions susceptibles d'être retenues. La valeur ajoutée est optimale sur des questions de type « Pourquoi?, Comment?, Pour qui?, et dans quels contextes ou circonstances? ».

Étape 2 du cycle de l'évaluation réaliste : élaborer la théorie de programme initiale

La théorie de programme initiale peut être élaborée en utilisant un ou plusieurs processus (figure 1).

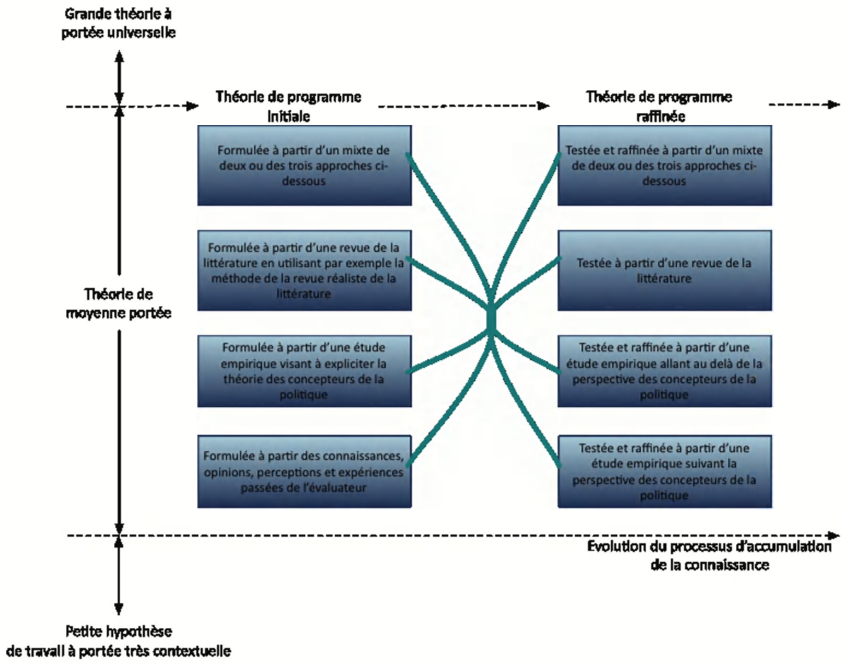


Figure 1 : Illustration des théories de moyenne portée et de différentes approches pour formuler et raffiner les théories de programmes (adapté de Merton (1968) et Marchal (et al. 2013))

Le choix de la stratégie pour formuler la théorie de programme initiale dépend de la nature de l'intervention, de l'étendue des connaissances de l'évaluateur ou de l'évaluatrice à propos du sujet à l'étude, des objectifs de l'évaluation, des ressources disponibles et de la nature des produits prioritaires attendus de l'évaluation. L'utilisation d'une combinaison des différentes approches est le modèle idéal.

Étape 3 du cycle de l'évaluation réaliste : choisir le devis de l'étude

L'évaluation réaliste n'impose aucun devis d'étude. Ce choix dépend essentiellement des questions à l'étude et des hypothèses qui constituent la théorie de programme initiale. Il est toutefois essentiel d'utiliser des devis qui permettent d'avoir des modèles d'inférence configurationelle et non des modèles d'inférence basés sur des variables (Van Belle *et al.* 2016). En rapport avec les questions d'évaluation, l'évaluation réaliste utilise la plupart du temps les méthodologies mixtes, combinant des méthodes quantitatives et

qualitatives à des degrés divers pour répondre aux questions d'étude (Ridde et Dagenais 2012).

Étape 4 du cycle de l'évaluation réaliste : collecter les données

La collecte des données se fait en fonction des méthodes choisies comme n'importe quelle évaluation. Il existe une forme particulière d'entretien qualitatif appelé « entretien réaliste ». Dans cette approche, le chercheur ou la chercheuse partage sa théorie avec les parties prenantes et leur demande leurs opinions sur les différentes hypothèses ainsi énoncées. Ce faisant, le chercheur ou la chercheuse donne sa théorie aux personnes interviewées qui partagent avec lui ou elle les éléments de leur théorie suivant le principe « Donne-moi ta théorie et je te donnerai la mienne ». Cette technique n'est pas une prescription absolue.

Étape 5 du cycle de l'évaluation réaliste : analyser les données

L'analyste réaliste identifie par argumentation inductive ou déductive les mécanismes à partir des données empiriques et formule par rétroduction les liens implicites entre les conditions du contexte et les mécanismes par lesquels les résultats observés sont susceptibles d'être expliqués.

La rétroduction (ou l'abduction) est un mode de raisonnement pour la découverte, distinct de l'induction et de la déduction. Elle part des résultats et des éléments contextuels observables pour formuler des explications plausibles. Le processus d'analyse réaliste peut ainsi prendre la forme suivante : au regard des résultats R, en présence des acteurs et actrices A et des éléments de contexte C1, C2 et C3, ces résultats ne peuvent être observés que s'il existe un raisonnement ou une réaction des acteurs et actrices (mécanisme sous-jacent M) à propos des ressources apportées par l'intervention I. L'identification de ces agrégats Intervention-Contexte-Acteurs-Mécanisme-Résultat ou *Outcome* (ICAMO) (Marchal *et al.*, 2018) constitue le travail essentiel de l'analyste réaliste.

La théorie de programme guide l'analyste réaliste qui veille aussi à rester ouvert-e aux nouvelles explications qui émergent des données. Les étapes clés dans le processus d'analyse sont les suivantes : 1) identification des résultats, 2) description de l'intervention en termes de contenu, de lieu, de temps et d'intensité, 3) identification des acteurs et actrices impliqués ou non dans la conception et dans la mise en œuvre, 4) identification des liens implicites entre les pratiques autour de l'intervention et les résultats observés (les mécanismes), 5) identification des caractéristiques du contexte

et de son influence différentielle sur le déclenchement des mécanismes et donc des résultats et 6) recherche active des explications alternatives. Ces étapes devraient permettre d'identifier les éléments des configurations ICAMO.

Pour l'analyse des données qualitatives, les critères de qualité comme l'analyse itérative, cumulative, adaptative et réflexive doivent être respectés (Teddlie et Tashakkori, 2009).

Étape 6 du cycle de l'évaluation réaliste : faire la synthèse

Les configurations sont ensuite synthétisées de manière narrative ou graphique. La formulation narrative peut adopter la structure de base : dans un contexte C, on observe un résultat R, si un mécanisme M est déclenché chez les acteurs et actrices A par l'intervention I.

Dans la synthèse, les mécanismes et leurs conditions d'exercice mis en évidence sont confrontés à la théorie de programme initiale qui est ainsi reformulée par addition des hypothèses pertinentes manquantes. Les hypothèses infirmées peuvent être reformulées de manière négative ou supprimées. Dans la démarche méthodologique, nous suggérons à cette étape de présenter dans la première colonne d'un tableau les hypothèses de la théorie de programme initiale et dans la seconde, la nouvelle formulation de ces hypothèses en mettant en évidence les altérations de raffinement apportées. Il s'agit d'une bonne pratique recommandée dans le rapportage de l'évaluation réaliste (Wong *et al.* 2014). La synthèse de ce cycle peut être rapportée sous forme de figure comme illustrée dans la figure 3 à la fin de ce chapitre.

Application de l'approche de l'évaluation réaliste

L'évaluation réaliste de l'adoption de la politique de gratuité de la césarienne dans les hôpitaux du Bénin peut nécessiter plusieurs cycles, chaque cycle expliquant de manière plus fine un ou plusieurs aspects des variations dans l'adoption de la politique d'un hôpital à l'autre. Les cycles subséquents peuvent aussi s'adapter au fur et à mesure aux questions qui émergent des cycles précédents. L'exemple rapporté ici porte sur un cycle.

Étape 1 du cycle de l'évaluation réaliste : la question de recherche

Pour l'évaluation de la mise en œuvre de la politique de gratuité de la

césarienne au Bénin, la question de recherche était la suivante : Comment, par qui et dans quelles circonstances la politique de gratuité de la césarienne au Bénin a-t-elle été adoptée dans les hôpitaux et pourquoi?

Étape 2 du cycle de l'évaluation réaliste : la théorie de programme initiale

Pour élaborer la théorie de programme initiale de la politique à l'étude, nous avons effectué une revue non systématique de la littérature qui a porté sur les théories, les cadres d'analyse et les modèles les plus à même de répondre seuls ou de manière complémentaire et cohérente à la question de recherche. La sélection des publications s'est faite suivant la stratégie du « *berry-picking* » (Booth 2008). Cette approche itérative et flexible commence par une question de recherche bibliographique. Le chercheur ou la chercheuse effectue une recherche documentaire, sélectionne une première série d'articles pertinents pour sa question, analyse les connaissances issues de ces publications, formule une autre question plus raffinée sur le même sujet, procède à une autre collecte de données et ainsi de suite, jusqu'à aboutir à une connaissance synthétique répondant de manière pragmatique à ses besoins de connaissance.

Nous avons donc commencé à sélectionner les publications dans la littérature sur les politiques d'exemption des frais de soins en santé maternelle en Afrique subsaharienne. Nous avons constaté un manque relatif de publications développant des théories, des cadres d'analyse ou des modèles pertinents (Ridde et Olivier de Sardan 2016). Nous avons décidé donc d'étendre notre recherche au domaine plus large de la mise en œuvre des politiques publiques et du *management*.

Il faut noter que cette évaluation a été réalisée dans le cadre de l'étude FEMHealth qui est une évaluation multi-pays, avec des parties prenantes situées à différents niveaux de la pyramide sanitaire. Nous estimions qu'une bonne théorie de programme initiale dans un tel contexte d'évaluation devrait :

- avoir une architecture de base assez complète pour permettre de structurer la collecte à différents niveaux de la pyramide sanitaire et auprès des différents groupes de parties prenantes
- être facile à comprendre et à communiquer afin qu'elle puisse structurer aisément les réflexions, les discussions et les connaissances

produites avec les multiples parties prenantes au niveau des pays.

Ce processus a abouti à la sélection de trois publications clés (Berman 1978, Matland 1995 et Elmore 1979) qui ont servi de base à l'élaboration de la théorie de programme initiale. Ces publications visaient explicitement à développer des modèles de mise en œuvre des politiques publiques et répondaient aux exigences définies plus haut liées au contexte de l'évaluation. La théorie de programme initiale qui en a résulté était formulée de manière narrative comme suit (encadré 1).

Encadré 1 : Théorie de programme initiale expliquant les variations dans l'adoption des politiques de santé d'une structure de mise en œuvre à l'autre

Un programme de santé est adopté au niveau opérationnel par les gestionnaires locaux qui sont responsables de l'appliquer et qui peuvent le faire de manière variable. Les facteurs personnels aussi bien que des facteurs contextuels locaux jouent un rôle, mais d'autres éléments au niveau de la politique, au niveau de sa traduction dans un programme et au niveau communautaire, sont également importants.

Il y a trois scénarios possibles dans l'adoption d'un programme par les gestionnaires locaux. Le premier scénario est la décision d'adopter et d'appliquer le programme sans modification. Ceci est fort probable si le programme est perçu comme une extension des objectifs locaux (alignement), c'est-à-dire s'il répond aux besoins locaux et fournit les ressources nécessaires. Le second scénario est l'adaptation (positive) au contexte local ; il est susceptible de se produire lorsque les gestionnaires locaux sont convaincus que le programme peut être amélioré. Cela nécessite aussi qu'ils et elles en aient la capacité et une marge de liberté pour ajuster le programme à la situation locale. Un tel ajustement peut mener à de meilleurs résultats dans la mise en œuvre de la politique. Le troisième scénario est le détournement du programme ; ceci peut se produire se le programme est perçu comme un moyen de créer des opportunités personnelles ou organisationnelles et si sa mise en œuvre est peu « soutenue » et que la contre-pression locale est faible. Le scénario réel final dépend de la pression locale, publique et politique (qui à son tour nécessite des représentants de la communauté bien informés sur la politique et ayant des canaux efficaces pour se faire entendre), et l'engagement personnel des gestionnaires locaux. Une mise en œuvre durable exige une institutionnalisation au niveau local.

Au niveau politique, l'engagement des gestionnaires et des bénéficiaires dans le processus décisionnel et la mesure dans laquelle les objectifs politiques et les groupes cibles sont clairement définis sont des facteurs influents.

Les facteurs au niveau du programme comprennent l'existence d'instructions opérationnelles claires (y compris l'identification des objectifs de la politique, les groupes cibles, les modalités pratiques de la mise en œuvre du remboursement, etc.). Si les problèmes du personnel de santé, tels qu'une compensation pour perte de revenus, une formation, une (re) distribution, etc., sont pris en main de façon appropriée, le programme aura plus de chances d'être adopté et appliqué.

Le passage d'un niveau du système à l'autre peut modifier la politique :

- Un bon consensus sur les objectifs, les groupes cibles et les modalités de la mise en œuvre (faible niveau d'ambiguïté et de conflit) peut faciliter le passage d'un niveau du système à l'autre si chaque niveau dispose des capacités et des ressources nécessaires.
- Un suivi et une capacité d'application faibles peuvent mener à la mise en œuvre symbolique s'il y a un sentiment profond de professionnalisme (dépendant de la volonté des professionnels locaux).
- S'il n'y a pas de consensus sur les objectifs, les groupes cibles et les modalités de la mise en œuvre, la politique peut être empêtrée dans des conflits de pouvoir ou transformée au niveau local.

Étape 3 du cycle de l'évaluation réaliste : le devis de l'étude

Nous avons choisi le devis d'étude de cas intégrés (Yin 2003) qui est bien adapté aux situations complexes (Marchal 2013). L'objectif de la phase de développement des outils et de collecte des données était de recueillir

de manière optimale des données empiriques susceptibles de permettre la formulation de conclusions de type configurationnel. Nous avons adopté une méthode mixte de type simultanée et imbriquée (*concurrent nested*) avec une composante quantitative plus petite intégrée dans une étude à dominante qualitative. Les composantes de la théorie de programme ont guidé ce développement d'outils.

La mise en œuvre de la politique de gratuité de la césarienne à l'échelle nationale a été considérée comme le cas d'étude. Le choix des unités d'analyse s'est fait de manière raisonnée pour permettre de « tester » les hypothèses exprimées par la théorie de programme initiale. Deux hôpitaux, dont les résultats dans la mise en œuvre de la politique sont contrastés (cas 1 et cas 2 sur la figure 1), ont été sélectionnés comme les unités d'analyse.

Étape 4 du cycle de l'évaluation réaliste : la collecte des données

Nous avons combiné une revue documentaire avec extraction de données quantitatives, des entretiens semi-structurés, une étude de coût et un suivi des flux financiers pour répondre à une série de questions intermédiaires qui visaient à contribuer à la réponse à la question centrale. Le tableau suivant présente les sources de données et les outils pour les études de cas.

Tableau 2. Sources de données, techniques de collecte et échantillon par unité d'analyse

Sources de données	Techniques	Cas 1 : Hôpital privé confessionnel	Cas 2 : Hôpital public
Gestionnaires locaux de la santé	Entretiens semi-structurés	4	4
Gestionnaires des hôpitaux et prestataires de soins	Entretiens semi-structurés	15	8
Représentants de la communauté dans les comités locaux de santé	Entretiens semi-structurés	6	3
Les femmes enceintes ciblées par la politique	Entretiens semi-structurés	6	6
	Enquêtes de sortie	28	17
Rapports sur la politique et différents documents comptables à l'Agence Nationale de Gratuité de la Césarienne et dans les structures de mise en œuvre	Suivi des flux financiers	1	1
Données du Système d'information sanitaire, prestataires de soins et gestionnaires locaux de la santé	Revue documentaire avec extraction de données pour étude de coûts	30	30
Environnement des hôpitaux de mise en œuvre	Observation non participante des pratiques autour de la politique	1	1

La collecte des données a été réalisée par une équipe multidisciplinaire de socio-anthropologues, d'économistes de la santé et de médecins. Elle a duré de mars 2012 à septembre 2012.

L'étude sur les perceptions, le vécu et les opinions des différentes parties prenantes à tous les niveaux a constitué un élément clé : elle a intégré les

observations préliminaires réalisées avec tous les autres outils de manière itérative, cumulative, adaptative et flexible.

Les interactions fréquentes au sein de l'équipe de collecte ont permis d'intégrer les données pertinentes produites par l'équipe multidisciplinaire avec les différentes techniques de collecte et d'adapter la collecte en conséquence au fur et à mesure.

Étapes 5 et 6 : l'analyse et la synthèse

Les données qualitatives ont été gérées par le logiciel NVIVO 10. Les mémos de collecte, les rapports des analyses préliminaires des autres techniques, ainsi que les documents de politiques, etc. ont été importés dans le logiciel pour constituer une base de données unique utilisée pour l'analyse. Les différentes composantes de la théorie initiale ont guidé la codification. L'analyse était réalisée de manière à permettre aussi l'addition de nouveaux codes de manière inductive.

Les analyses préliminaires effectuées dans les différentes disciplines impliquées dans l'étude ont été importantes pour l'analyse réaliste proprement dite.

Dans la synthèse, les mécanismes et les hypothèses ainsi mis en évidence ont été confrontés à la théorie initiale, qui a été reformulée par addition des hypothèses pertinentes.

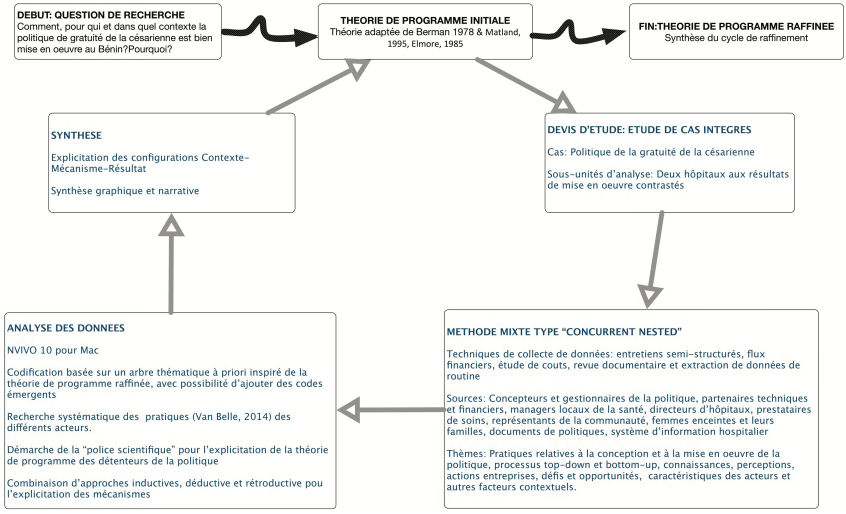


Figure 3. Cycle de l'évaluation réaliste de la mise en œuvre de la politique de gratuité de la césarienne au Bénin

Quelques résultats

Dans cette section, nous présentons de manière illustrative quelques résultats issus des différentes étapes méthodologiques. Il ne s'agit pas de résultats exhaustifs sur le sujet à l'étude, mais d'exemples pour illustrer le processus et la manière dont les résultats d'une évaluation réaliste pourraient être formulés.

Les pratiques autour de l'adoption de la politique par les gestionnaires des hôpitaux

Dans l'hôpital privé confessionnel

Dans cet hôpital, sur les huit éléments du paquet de la politique, seul l'examen postopératoire était entièrement exempté. La rémunération médiane payée par les usagères pour une césarienne après le lancement de la police s'élevait à 48 $US [min. 45 $US ; max 135 $US] au lieu d'être totalement gratuit comme prévu par la Politique promue par l'État.

Cet hôpital est la propriété d'une organisation religieuse qui en possède les infrastructures, qui nomme les gestionnaires, recrute le personnel, valide le budget, mobilise des ressources extérieures pour le fonctionnement de l'hôpital au besoin et contrôle la gestion. L'hôpital fonctionne essentiellement sur la base de ressources internes mobilisées à travers le recouvrement des coûts par paiement direct. Les interventions de l'État dans le fonctionnement de cet hôpital sont minimes. Le Comité de Santé – instance locale d'implication de la communauté dans la gouvernance du système local de santé – n'intervient quasiment jamais dans cet hôpital. Les populations sont peu informées à propos de la politique dans cet hôpital et n'ont aucun canal officiel pour s'informer ou exprimer leurs plaintes.

Le décret n° 2008-730 du 22 décembre 2008, portant institution de la Politique, a accrédité cet hôpital pour la mise en œuvre de la Politique dès son lancement en avril 2009. Cependant, les gestionnaires de cet hôpital n'ont adopté et n'ont commencé à offrir effectivement la césarienne gratuite qu'en septembre 2009. Les gestionnaires ont mentionné qu'ils ne faisaient pas confiance au gouvernement pour effectuer les remboursements en temps opportun, tel qu'indiqué dans les documents de politique officiels. Les gestionnaires ont justifié cette méfiance par leurs expériences antérieures

avec l'État qui continue à leur devoir, depuis plusieurs années, les ressources financières exemptées aux bénéficiaires de la couverture sanitaire partielle accordée aux fonctionnaires. Les gestionnaires ont commencé à mettre en œuvre l'exemption de paiement pour les césariennes lorsqu'ils et elles ont constaté que les autres hôpitaux privés mettant en œuvre la politique ont commencé à recevoir effectivement les remboursements de manière régulière et sans trop de retard.

L'étude des coûts a montré que le coût de production médian (y compris tous les coûts directs) d'une césarienne dans cet hôpital confessionnel était de 44 $US [min 37 $US ; max 69 $US]. Les frais facturés aux usagères avant la politique étaient de 163 $US. Notre outil de suivi des flux financiers a montré que jusqu'en décembre 2011, cet hôpital était remboursé pour toutes les césariennes effectuées. Les gestionnaires ont perçu que le montant fixe de 178 $US était insuffisant et injuste, estimant que le taux de remboursement pour les hôpitaux n'appartenant pas à l'État devait être supérieur à celui des hôpitaux appartenant à l'État où les subventions publiques couvrent de nombreux coûts de fonctionnement. Les gestionnaires ont donc décidé de facturer systématiquement 35 $US aux usagères par césarienne en plus du remboursement reçu de l'État.

Dans l'hôpital public

Sur les huit éléments inclus dans le paquet de la césarienne gratuite, cinq ont été entièrement exemptés de frais de soins dans cet hôpital. Les frais encore à la charge des usagères s'élevaient à 6,22 $US [0,1 $US min ; max 53,3 $US].

Cet hôpital est la propriété de l'État qui l'a construit et l'a équipé avec des ressources publiques. L'État y paye une partie importante du personnel et nomme les gestionnaires. La population est mieux informée sur la politique. Les usagers et usagères utilisent plusieurs canaux comme les radios communautaires pour s'informer sur la politique ou se plaindre au besoin.

Les gestionnaires de cet hôpital ont adopté la politique rapidement et ont commencé à la mettre en œuvre moins d'une semaine après le lancement officiel. Le directeur de l'hôpital a participé à l'atelier technique qui a élaboré des options stratégiques dans la phase de développement de la politique. Il a perçu la politique comme un soutien à sa vision d'augmenter l'accès financier aux services pour les utilisateurs et utilisatrices dans son hôpital. Il a pris des mesures actives pour faciliter sa mise en œuvre réussie.

Dans cet hôpital, les frais de remboursement ont été perçus comme couvrant largement les coûts de la césarienne pour l'hôpital. Les gestionnaires qui sont des fonctionnaires de l'État se sont sentis obligés de commencer à mettre en œuvre la politique, comme l'a expliqué un gestionnaire :

> La population a été largement informée que la césarienne est gratuite et il est difficile de dire maintenant, moi, je ne vais pas commencer à mettre en œuvre la politique. Une femme ne peut pas venir et vous lui dites, nous n'exécutons pas de césarienne gratuite aujourd'hui, puisque le directeur provincial de la santé sait que la césarienne est gratuite. Si une femme se plaint que l'établissement lui a demandé de payer, en tant que gestionnaire, vous devrez expliquer pourquoi. Gestionnaire, Hôpital public, 2012

Quelques configurations causales autour de l'adoption dérivée des résultats

Notre analyse des pratiques énoncées dans l'hôpital privé confessionnel nous indique que, dans ce contexte, les gestionnaires ne sont pas dans une relation administrative hiérarchique avec l'État, car les pressions issues des populations sont trop faibles ou absentes. Dans ce contexte, l'administration publique est perçue comme lente, peu crédible à propos du respect de ses engagements financiers et suscite une attitude de méfiance alors que la mobilisation des ressources issues des prestations des soins est essentielle pour le fonctionnement de l'hôpital. Le paiement direct est systématique, fortement ancré dans la culture organisationnelle et considéré comme une mesure capitale de survie organisationnelle. Dans ce contexte, la politique d'exemption de paiement des frais de soins (Intervention) est perçue par les gestionnaires d'hôpitaux (Acteurs) comme une menace pour la survie des hôpitaux (Mécanisme) et de fait un scénario sub-optimal dans l'adoption de la politique (Résultat).

La bonne adoption (Résultat) dans ce contexte nécessite que la politique soit perçue comme protectrice des intérêts financiers des hôpitaux ou au mieux qu'elle apporte une plus-value financière aux hôpitaux (Mécanisme), par l'apport de ressources plus facilement recouvrables, plus abondantes ou plus régulières (Intervention).

Dans le contexte de l'hôpital public, notre analyse montre que les

gestionnaires sont relativement plus susceptibles de se conformer aux injonctions administratives hiérarchiques venant de l'État. La participation communautaire à la gestion de l'hôpital est relativement plus importante avec des populations bien informées sur la politique et disposant de canaux pour exprimer leurs plaintes. Cette double pression, administrative descendante et communautaire ascendante, sur les gestionnaires des hôpitaux (Acteurs) génère une coercition (Mécanisme) qui induit une meilleure adoption de la politique (Résultat).

Les investissements structurels de l'État avant l'initiation de la politique dans l'hôpital public sont plus importants que dans l'hôpital privé (Contexte). Dans ce contexte, le remboursement d'un même montant forfaitaire par acte (Intervention) génère un sentiment d'injustice (Mécanisme) auprès des structures n'appartenant pas à l'État (Acteurs) qui tendent de fait à compenser cette injustice en percevant des frais supplémentaires auprès des usagers et usagères, n'offrant donc en fait qu'une partie des services gratuitement (Résultat).

Synthèse et discussion : les théories autour de l'adoption de la politique

À travers les configurations causales formulées plus haut, on constate que l'adoption dépend de la confiance entre les parties prenantes et le promoteur de la politique d'exemption, par exemple l'État. Cette confiance est d'autant plus nécessaire si la crédibilité du promoteur est remise en cause sur la base d'expériences antérieures. Les dispositions à prendre pour rassurer les gestionnaires des hôpitaux incluent l'apport à temps de ressources adaptées à l'offre du service, tout au moins en compensation suffisante pour la perte de ressources générée par l'exemption de paiement.

Cette explication est compatible avec la théorie du comportement planifié qui stipule qu'il est peu probable que les individus développent une forte intention d'agir et se comportent d'une certaine manière s'ils pensent qu'ils n'ont pas les ressources et les opportunités pour y parvenir, même s'ils ont une attitude favorable envers le comportement en question et même s'ils estiment que les membres de leurs familles ou leurs congénères approuveraient le comportement (normes subjectives) (Ajzen et Fishbein 1980). Ainsi, même si la volonté politique et les injonctions administratives insufflent la formation de nouvelles normes et de nouvelles structures de

référence, il est peu probable que l'adoption de la politique soit une réalité si elle n'a pas des caractéristiques requises et si elle n'apporte pas suffisamment de ressources garanties présentes et futures pour renforcer le contrôle perçu des acteurs et actrices sur sa mise en œuvre.

La perception de la justice dans la répartition des ressources est capitale dans le cas où les structures de mise en œuvre font l'objet de droits de propriété différents et sont sujettes à des financements variables. Un sentiment d'injustice est susceptible de se produire si la politique ne définit pas clairement les coûts réels engagés dans la production de chaque acte par type de formation sanitaire et si elle ne définit pas de manière explicite la marge excédentaire accordée à chaque unité de mise en œuvre. Ce besoin d'équité perçu dans l'adoption n'est présent ni dans la théorie du comportement planifié (Ajzen et Fishbein 1980), ni dans le modèle de diffusion des innovations de Rogers (1995), ni dans le modèle d'acceptation des technologies (Davis 1989). L'influence de ce sentiment d'injustice sur l'adoption de la politique peut être analysée comme une pression isomorphique mimétique (Zucker 1988), dans laquelle les acteurs et actrices des structures de santé n'appartenant pas à l'État en apprennent sur les avantages tirés de la politique par les structures de santé appartenant à l'État et recherchent à travers leurs comportements des bénéfices similaires.

Cette analyse nous a amenés à reformuler la théorie de programme initiale (encadré 2).

Encadré 2 : Théorie de programme raffinée

Un programme de santé est adopté au niveau opérationnel par les gestionnaires locaux de la santé qui sont responsables de l'appliquer et qui peuvent le faire de manière variable. Les facteurs personnels aussi bien que des facteurs contextuels locaux jouent un rôle, mais d'autres éléments au niveau de la politique, sa traduction dans un programme et au niveau communautaire, sont également importants.

Il y a quatre scénarios possibles dans l'adoption d'un programme par les gestionnaires locaux. Le premier scénario est la décision d'adopter et d'appliquer le programme sans modification. Ceci est fort probable si le programme est perçu comme une extension des objectifs locaux (alignement), c'est-à-dire s'il répond aux besoins locaux et fournit les ressources nécessaires. Le second scénario est l'adaptation (positive) au contexte local. Il est susceptible de se produire lorsque les gestionnaires locaux sont convaincus que le programme peut être amélioré. Cela nécessite aussi qu'ils et elles en aient la capacité et une marge de liberté pour ajuster le programme à la situation locale. Un tel ajustement peut mener à de meilleurs résultats dans la mise en œuvre de la politique. Le troisième scénario est le détournement du programme ; ceci peut se produire si le programme est perçu comme un moyen de créer des opportunités personnelles ou organisationnelles et si sa mise en œuvre est peu « soutenue » et que la contre-pression locale est faible. Un quatrième scénario possible est la décision d'adopter et d'appliquer une version réduite du programme dans laquelle une partie seulement des services est offerte gratuitement. Ce scénario est susceptible de se produire même si les ressources apportées par la politique sont suffisantes. Il est plus probable si les gestionnaires se sentent défavorisés dans l'allocation des ressources apportées par la politique par rapport à d'autres structures de mise en œuvre. Le scénario réel final dépend de la pression locale, publique et politique (qui à son tour nécessite des représentant-e-s de la communauté bien informés sur la politique et ayant des canaux efficaces pour se faire entendre), la confiance des gestionnaires des structures de mise en œuvre dans l'engagement du promoteur de la politique et l'engagement personnel des gestionnaires locaux pour le succès de la politique. Une mise en œuvre durable exige une institutionnalisation au niveau local.

Au niveau politique, l'engagement des gestionnaires et des bénéficiaires dans le processus décisionnel et la mesure dans laquelle les objectifs politiques et les groupes cibles sont clairement définis sont des facteurs influents.

Les facteurs au niveau du programme comprennent l'existence d'instructions opérationnelles claires (y compris l'identification des objectifs de la politique, les groupes cibles, les modalités pratiques de la mise en œuvre du remboursement, etc.). Si les problèmes du personnel de santé, tels qu'une compensation pour perte de revenus, une formation, une (re) distribution, etc., sont pris en main de façon appropriée, le programme aura plus de chances d'être adopté et appliqué.

Le passage d'une phase à l'autre peut modifier la politique :
– Un bon consensus sur les objectifs, les groupes cibles et les modalités de la mise en œuvre (faible niveau d'ambiguïté et de conflit) peut faciliter le passage d'une phase à l'autre si chaque niveau dispose des capacités et des ressources nécessaires.
– Un suivi et une capacité d'application faibles peuvent mener à la mise en œuvre symbolique s'il y a un sentiment profond de professionnalisme (dépendant de la volonté des professionnels locaux).
– S'il n'y a pas de consensus sur les objectifs, les groupes cibles et les modalités de la mise en œuvre, la politique peut être empêtrée dans des conflits de pouvoir ou transformée au niveau local.

Dans cette théorie de programme raffinée, de nouvelles hypothèses sont ajoutées à la théorie de programme initiale par exemple à propos du rôle du sentiment d'injustice et de la confiance dans l'adoption de la politique. Le raffinement a porté aussi sur une modification d'une hypothèse existante. Par exemple, le nombre de scénario d'adoption possibles est passé de trois à quatre.

Cette théorie de programme raffinée apporte une réponse plausible à la question d'évaluation. Les décideuses et décideurs politique peuvent s'en inspirer de manière pratique pour améliorer la conception et la mise en œuvre de cette politique. Cette théorie de programme peut aussi permettre de mieux concevoir d'autres politiques similaires.

Cette théorie de programme reste toutefois une théorie de programme qui ne vise d'aucune manière à être universelle ou transférable à tous les contextes ou à tous les types d'intervention. Elle est par nature susceptible d'être améliorée au fil de cycles d'évaluation subséquentes.

Leçons apprises et discussion sur les enjeux de la mise en œuvre de l'évaluation

- Il faut une bonne préparation. Même si cette phase n'a pas forcément d'aspects fondamentalement différents de la préparation des autres formes d'évaluation, elle doit permettre de rassembler autant de données que possible sur les partenaires locaux de recherche, l'intervention à l'évaluation, les parties prenantes autour de l'intervention, les contextes de développement et de mise en œuvre. Nous suggérons à cette étape de distinguer clairement les deux situations suivantes :
    - le choix est fait de développer la théorie de programme initiale à partir de celle des concepteurs et conceptrices de l'intervention. Dans ce cas, il n'est pas indispensable pour l'évaluateur ou l'évaluatrice de finaliser une théorie de programme initiale avant de démarrer les activités de terrain. Il ou elle doit toutefois avoir une connaissance théorique et méthodologique suffisante pour expliciter la théorie de programme initiale des concepteurs et conceptrices dans une logique réaliste.
    - le choix est fait de développer la théorie de programme initiale

indépendamment ou en complément de celle des concepteurs et conceptrices de l'intervention. Dans ce cas, il est essentiel de consacrer du temps à des revues de littérature.

- Il faut maîtriser une définition claire du concept de « mécanisme causal » au départ de l'évaluation. Il s'agit de l'un des défis majeurs pour comprendre l'évaluation réaliste.
- La distinction entre mécanisme, contexte et résultats peut parfois devenir un casse-tête et une source de confusion. Nous avons appris que la définition la plus opérationnelle est celle des « raisonnements et des réactions des acteurs et actrices à propos des ressources apportées par l'intervention ». En d'autres termes, les mécanismes sont cachés dans les perceptions, les opinions, les logiques implicites ou explicites des acteurs et actrices qui sous-tendent leurs attitudes, leurs actions vis-à-vis des ressources apportées par l'intervention, ce qui en détermine les résultats (Pawson et Tilley 1997; Lacouture *et al.* 2015). Ces mécanismes peuvent être explicitement formulés dans les discours des acteurs et actrices ou faire l'objet d'une rétroaction pour être identifiés.
- Il est essentiel de centrer la collecte de données sur l'approche configurationelle : il faut parfois des outils spécifiques pour documenter le contexte, la mise en œuvre de l'intervention, les résultats de l'intervention et les mécanismes, le tout construit sur la théorie de programme initiale. Pendant l'analyse, il est essentiel de garder une vision intégrée de ces composantes en triangulant les données afin d'identifier des configurations dans lesquelles l'intervention, les acteurs et actrices, les mécanismes, les éléments contextuels et les résultats sont en interaction.
- Le raffinement de la théorie de programme est un processus qui n'aboutit jamais. L'objectif final d'un cycle n'est pas de produire une théorie de programme universelle, mais utile en produisant des connaissances pertinentes pour les parties prenantes ou les chercheuses et chercheurs. Ce regard pragmatique sur la finalité d'un cycle d'évaluation réaliste est essentiel pour guider les efforts et éviter de s'embourber dans un processus sans fin.

Références clés

Pawson, R. et Tilley, N. (1997). *Realistic evaluation*. Londres : Sage Publications.

Les auteurs de ce livre sont les pères fondateurs de l'évaluation réaliste. Dans ce livre, ils en expliquent longuement les fondements et illustrent les concepts clés par des exemples en s'appuyant sur des programmes sociaux de différents domaines dans les pays développés.

Marchal, B., Van Belle, S., Van Olmen, J., Hoerée, T. et Kegels, G. (2012). Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research. *Evaluation*, 18(2), 192–212.

Cette publication fait une revue de l'utilisation de l'évaluation réaliste dans le cadre de la recherche sur les systèmes de santé. Elle décrit entre autres les différents modèles d'utilisation, les confusions conceptuelles et les défis méthodologiques observés. Elle propose le cycle de l'évaluation réaliste et en décrit les différentes étapes pour réduire les confusions et faciliter la structuration de la démarche méthodologique.

Lacouture, A., Breton, E., Guichard, A. et Ridde, V. (2015). The concept of mechanism from a realist approach: a scoping review to facilitate its operationalization in public health program evaluation. *Implementation Science*, 10, 153-162.

Cette publication fait une synthèse des différentes définitions opérationnelles du concept de « mécanisme » dans l'évaluation des politiques de santé publique. Elle est utile pour mieux comprendre ce concept clé et peut aider à s'affranchir des confusions possibles entre les concepts de « mécanisme », « contexte » et « résultats ».

Site http://www.ramesesproject.org/

Ce site comporte les produits des projets Realist And Meta-narrative Evidence Syntheses : Evolving Standards (RAMESES) I et II. RAMESES I a développé des standards de qualité et de publication ainsi que les matériels de formation pour les revues réalistes ainsi que les approches y afférentes des revues méta-narratives. RAMESES II a développé des standards de qualité et de publication ainsi que les matériels de formation pour l'évaluation réaliste. Ce site est en anglais.

Références

Ajzen, I. et Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ : Prentice-Hall.

Audi, R. (1999). *The Cambridge dictionary of philosophy* (2e éd.). Cambridge, NY: Cambridge University Press.

Berman, P. (1978). *The study of macro and micro implementation of social policy*. Santa Monica, CA : RAND Corporation.

Booth, A. (2008). Unpacking your literature search toolbox: on search styles and tactics. *Health Information et Libraries Journal*, 25(4), 313–317.

Centre de Recherche en Reproduction humaine et en Démographie. (2014). *Rapport de synthèse : l'évaluation de la politique de gratuité de la césarienne dans cinq zones sanitaires au Bénin*. Cotonou, Bénin: University of Aberdeen.

Dalkin, S. M., Greenhalgh, J., Jones, D., Cunningham, B. et Lhussier, M. (2015). What's in a mechanism?: Development of a key concept in realist evaluation. *Implementation Science*, 10, 49-55.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *Management Information Systems Quarterly*, 13, 319–340.

Elmore, F. R. (1979). Backward Mapping : Implementation Research and Policy Decisions. *Polit Sci Q*, 94, 601-16.

Goodnough, L. T., Riddell, J., Lazarus, H., Chafel, T. L., Prince, G., Hendrix, D., et Yomtovian, R. (1993). Prevalence of platelet transfusion reactions before and after implementation of leukocyte-depleted platelet concentrates by filtration. *Vox Sanguinis*, 65(2), 103–107.

Lacouture, A., Breton, E., Guichard, A. et Ridde, V. (2015). The concept of mechanism from a realist approach: a scoping review to facilitate its operationalization in public health program evaluation. *Implementation Science*, 10, 153-162.

Marchal, B., Kegels, G. et Van Belle, S. (2018). Realist evaluation in health policy and systems research: theory incarnate. Dans N. Emmel, J. Greenhalgh, A. Manzano, M. Monaghan et S. Dalkin (dir.), *Doing Realist Research* (édition 2018, volume 1). Londres, Angleterre: Sage Publications.

Marchal, B., Van Belle, S., De Brouwere, V. et Witter, S. (2013). Studying complex interventions: reflections from the FEMHealth project on evaluating fee exemption policies in West Africa and Morocco. *BMC Health Services Research*, 13, 469-477.

Marchal, B., Van Belle, S., Van Olmen, J., HoerÇe, T. et Kegels, G. (2012). Is realist evaluation keeping its promise?: A review of published empirical studies in the field of health systems research. *Evaluation*, 18(2), 192–212.

Marchal, B., Van der Veken, K., Essolbi, A., Dossou, J. P., Richard, F. et Van Belle, S. (2013). *Methodological reflections on using realist evaluation in a study of fee exemption policies in West Africa and Morocco*. Aberdeen : University of Aberdeen.

Matland, R. E. (1995). Synthesizing the implementation literature: The ambiguity-conflict model of policy implementation. *Journal of Public Administration Research and Theory*, 5(2), 145-174.

Merton, R. K. (1968). *Social theory and social culture*. New York : The Free Press.

Pawson, R. (2006). *Evidence-based policy: a realist perspective*. London; Thousand Oaks, Calif. : SAGE Publications.

Pawson, R. et Tilley, N. (1997). *Realistic evaluation*. Londres : SAGE Publications.

Ridde, V. et Dagenais, C. (2012). *Approches et pratiques en évaluation de programmes*. Montréal : Presses de l'Université de Montréal.

Ridde, V. et Olivier de Sardan J. P. (2016). The implementation of public health interventions in Africa: a neglected strategic theme. *Médecine et Santé Tropicales*, 27(1), 6-9.

Ridde, V. et Olivier de Sardan J. P. (2017). The implementation of public health interventions in Africa: a neglected strategic theme. *Médecine et Santé Tropicales*, 27(1), 6–9.

Rogers, E. M. (1995). *Diffusion of innovations* (4e éd.). New York : Free Press.

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R. et Befani, B. (2012). *Broadening the range of designs and methods for impact evaluations*. Londres : Department for International Development.

Teddlie, C. et Tashakkori, A. (2009). *Foundations of mixed methods research*:

integrating quantitative and qualitative approaches in the social and behavioral sciences. Londres : SAGE Publications

Van Belle, S., Wong, G., Westhorp, G., Pearson, M., Emmel, N., Manzano, A. et Marchal, B. (2016). Can 'realist' randomised controlled trials be genuinely realist?. *Trials*, 17, 313-318.

Westhorp, G. (2014). *Realist impact evaluation: An introduction*. Londres : A method lab.

Westhorp, G., Prins, E., Kusters, C., Hultink, M., Guijt, I. et Brouwers, J. (2011). Realist evaluation: an overview. Report from an expert seminar with Dr. Gill Westhorp. Wageningen: Wageningen UR Centre for Development Innovation.

Wong, G., Greenhalgh, T., Westhorp, G. et Pawson, R. (2014). *Development of methodological guidance, publication standards and training materials for realist and meta-narrative reviews: the* RAMESES *(Realist And Meta-narrative Evidence Syntheses – Evolving Standards) project. Health Services and Delivery Research, No. 2.30*. Southampton, UK : NIHR Journals Library.

Yin, R. (2003). *Case study research. Design and methods*. Londres : Sage Publications.

Zucker, L. G. (1988). *Institutional patterns and organizations : culture and environment*. Cambridge, MA : Ballinger Publishing.

Résumé / Overzicht / Abstract / Resumen

L'évaluation réaliste (ER) explicite les mécanismes causaux des résultats d'une intervention dans différents contextes. En s'appuyant sur l'évaluation de la mise en œuvre d'une politique de santé en Afrique de l'Ouest, ce chapitre illustre les différentes étapes du cycle de l'ER. À la première étape, les questions de recherche sont formulées. La valeur ajoutée de l'approche de l'ER est optimale sur des questions de type « Pourquoi? Comment? Pour qui? et dans quels contextes une intervention marche ou non? ». La seconde étape est la formulation de la théorie de programme initiale qui dépend de la nature de l'intervention, les objectifs d'évaluation, la nature des produits prioritaires attendus de l'évaluation, les ressources disponibles et la familiarité de l'évaluateur avec l'objet d'évaluation. La collecte des données peut se baser sur une forme particulière d'entretien qualitatif appelé «

entretien réaliste ». En pratique, la collecte de données doit s'adapter aux informateurs clés, à leurs perceptions et attentes vis-à-vis des évaluateurs et évaluatrices pour éviter la tendance à la confirmation systématique de la théorie de l'évaluateur par les interviewés. L'analyste réaliste identifie activement les configurations « Intervention-Contexte-Acteur-Mécanisme-Résultat ». En pratique le raisonnement par rétroduction est utilisé pour expliciter les mécanismes. Un tableau permettant de suivre les changements apportés à la théorie de programme initiale facilite la dernière étape de la synthèse et du rapport de l'ER.

***

Realistische evaluatie (RE) verklaart de causale mechanismen die de resultaten van een interventie verklaren in verschillende contexten. Op basis van de evaluatie van de implementatie van een gezondheidsbeleid in West-Afrika illustreert dit hoofdstuk de verschillende stadia van de RE-cyclus. In de eerste fase worden de onderzoeksvragen geformuleerd. De toegevoegde waarde van de RE-benadering is optimaal voor vragen als « waarom, hoe, voor wie, en in welke context werkt een interventie of niet? ». De tweede stap is de formulering van de initiële programmatheorie, die afhangt van de natuur van de interventie, de evaluatiedoelstellingen, de aard van de outputs die van de evaluatie worden verwacht, de beschikbare middelen en de bekendheid van de beoordelaar met het doel van de evaluatie. Voor het verzamelen van gegevens kan elke methode gebruikt worden? Een specifieke vorm van interviews, genaamd « realistisch interview » wordt vaak gebruikt. In de praktijk moet het verzamelen van gegevens zich aanpassen aan de informanten, hun percepties en verwachtingen ten opzichte van de beoordelaars om de neiging tot systematische bevestiging van de initiële theorie door de geïnterviewden te vermijden. De realistische analist identificeert vervolgens de configuraties 'Interventie-context-actor-mechanisme-resultaat'. In de praktijk wordt retroductie gebruikt om de mechanismen te indetificeren. Een tabel om wijzigingen in de oorspronkelijke programmatheorie bij te houden, vergemakkelijkt de laatste stap van de synthese en rapportage van de RE.

***

Realist Evaluation (RE) identifies the causal mechanisms explaining the results of a given intervention in different contexts. Based on the evaluation of the implementation of a health policy in West Africa, this chapter

illustrates the different stages of the RE cycle. In the first stage, the research question is formulated. The added value of the RE approach is optimal on questions such as « why, how, for whom, and in what contexts does an intervention work or not? ». The second step is the formulation of the initial program theory, which depends on the nature of the intervention being studied, the objectives, and the nature of the main outputs expected from the evaluation, the resources available and the evaluator's familiarity with the object of evaluation The Data collection is method neutral and can be based on a particular form of qualitative interview called « realist interview ». In practice, data collection must adapt to key informants, their perceptions and expectations vis-à-vis the evaluators to avoid the tendency for systematic confirmation of the evaluator theory by the interviewees. The realist analyst actively identifies the « Intervention-Context-Actor-Mechanism-Outcome » configurations. In practice, retroductive analysis is used to explicit mechanisms that are implicit by nature. A table to track changes to the initial program theory facilitates the final step of the synthesis and the reporting of the RE.

***

La evaluación realista (ER) explica los mecanismos causales de los resultados de la intervención en diferentes contextos. Basado en la evaluación de la implementación de políticas de salud en África Occidental, este capítulo ilustra las diferentes etapas del ciclo de la ER. En el primer paso, se formulan las preguntas de la investigación. El valor añadido de la evaluación realista es óptimo para responder a preguntas como « ¿Por qué? ¿Cómo? ¿Para quién? y ¿en qué contextos funciona o no una intervención? ». El segundo paso es la formulación de la teoría del programa inicial, que depende de la naturaleza de la intervención, los objetivos de la evaluación, la naturaleza de los productos prioritarios principales esperados de la evaluación, los recursos disponibles y la familiaridad del evaluador con el tema de la evaluación. La recolección de datos puede basarse en una forma particular de entrevista cualitativa llamada « entrevista realista ». En la práctica, la recolección de datos debe adaptarse a los informantes clave, a sus percepciones y expectativas de los evaluadores para evitar la tendencia de los entrevistados a confirmar sistemáticamente la teoría del evaluador. El analista realista identifica activamente las configuraciones « Intervención-Contexto-Actor-Mecanismo-Resultado ». En la práctica, el razonamiento retroductivo se utiliza para explicar los mecanismos. Una tabla para el

seguimiento de los cambios en la teoría original del programa facilita el paso final de la síntesis y el informe de la ER.

***

Jean-Paul Dossou est médecin de santé publique, chercheur en politiques et systèmes de santé, sur des questions de santé sexuelle et reproductive et de couverture sanitaire universelle en Afrique de l'Ouest et de l'Est. Il forge passionnément ses compétences théoriques et pratiques en complexité et en méthodologies réalistes depuis 2012 en travaillant étroitement avec des chercheuses et chercheurs réalistes de référence mondiale, dont le Pr Bruno Marchal.

Bruno Marchal est chef de l'unité « Organisation des Services de Santé » au Département de Santé publique à l'Institut de Médecine Tropicale d'Anvers en Belgique. Ses travaux de recherche et d'enseignement portent principalement sur la gestion stratégique des services de santé, ainsi que sur des méthodologies de recherche adaptées aux questions complexes en santé. Il est l'auteur des publications les plus citées sur l'évaluation réaliste et est en première ligne dans le développement de l'évaluation réaliste dans la santé mondiale.

Citation

Jean-Paul Dossou et Bruno Marchal. (2019). L'évaluation réaliste. L'exemple de l'adoption d'une politique publique de santé au Bénin. In *Évaluation des interventions de santé mondiale. Méthodes avancées*. Sous la direction de Valéry Ridde et Christian Dagenais, pp. 433-462. Québec : Éditions science et bien commun et Marseille : IRD Éditions.

À propos de la maison d'édition

Les Éditions science et bien commun sont une branche de l'Association science et bien commun (ASBC), un organisme sans but lucratif enregistré au Québec depuis juillet 2011.

L'Association science et bien commun

L'ASBC a comme mission de stimuler la vigilance et l'action pour une science ouverte au service du bien commun. À cette fin, elle s'emploie à :

- Défendre et promouvoir une vision des sciences au service du bien commun;
- Colliger, analyser, produire et diffuser de l'information sur la science et sur ses rapports avec la société;
- Soutenir, promouvoir ou organiser des expériences de démocratisation des sciences;
- Organiser des expériences de débat public sur diverses facettes des sciences;
- Mettre en place des expériences de rencontre entre le monde scientifique et d'autres sphères sociales (ex. le milieu artistique, le milieu politique, etc.);
- Offrir un service d'orientation des groupes de la société civile dans le monde universitaire;
- Offrir, sous réserve de la Loi sur l'enseignement privé (L. R. Q., c. E-9.1) et de ses règlements, des formations sur la responsabilité sociale, la science avec les citoyens et l'éthique des sciences.

Sur son site Web se trouvent de nombreuses informations sur ses activités et ses publications. Il est possible de devenir membre de l'Association science et bien commun en payant un tarif modeste.

Pour plus d'information, écrire à info @ scienceetbiencommun.org, s'abonner à son compte Twitter @ScienceBienComm ou à sa page Facebook : https://www.facebook.com/scienceetbiencommun

Les Éditions science et bien commun

Un projet éditorial novateur dont les principales valeurs sont :

- la publication numérique en libre accès, en plus des autres formats
- la pluridisciplinarité, dans la mesure du possible
- le plurilinguisme qui encourage à publier en plusieurs langues, notamment dans des langues nationales africaines ou en créole, en plus du français
- l'internationalisation, qui conduit à vouloir rassembler des auteurs et auteures de différents pays ou à écrire en ayant à l'esprit un public issu de différents pays, de différentes cultures
- mais surtout la justice cognitive :
    - chaque livre collectif, même s'il s'agit des actes d'un colloque, devrait aspirer à la parité entre femmes et hommes, entre juniors et seniors, entre auteurs et auteures issues du Nord et issues du Sud (des Suds); en tout cas, tous les livres devront éviter un déséquilibre flagrant entre ces points de vue;
    - chaque livre, même rédigé par une seule personne, devrait s'efforcer d'inclure des références à la fois aux pays du Nord et aux pays des Suds, dans ses thèmes ou dans sa bibliographie;
    - chaque livre devrait viser l'accessibilité et la « lisibilité », réduisant au maximum le jargon, même s'il est à vocation scientifique et évalué par les pairs.

Le catalogue

Le catalogue des Éditions science et bien commun (ESBC) est composé de livres qui respectent les valeurs et principes des ÉSBC énoncés ci-dessus :

- Des ouvrages scientifiques (livres collectifs de toutes sortes ou monographies) qui peuvent être des manuscrits inédits originaux, issus de thèses, de mémoires, de colloques, de séminaires ou de projets de recherche, des rééditions numériques ou des manuels universitaires. Les manuscrits inédits seront évalués par les pairs de manière ouverte, sauf si les auteurs ne le souhaitent pas (voir le point de l'évaluation ci-dessus).

- Des ouvrages de science citoyenne ou participative, de vulgarisation scientifique ou qui présentent des savoirs locaux et patrimoniaux, dont le but est de rendre des savoirs accessibles au plus grand nombre.
- Des essais portant sur les sciences et les politiques scientifiques (en études sociales des sciences ou en éthique des sciences, par exemple).
- Des anthologies de textes déjà publiés, mais non accessibles sur le web, dans une langue autre que le français ou qui ne sont pas en libre accès, mais d'un intérêt scientifique, intellectuel ou patrimonial démontré.
- Des manuels scolaires ou des livres éducatifs pour enfants

Pour l'accès libre et universel, par le biais du numérique, à des livres scientifiques publiés par des auteures et auteurs de pays des Suds et du Nord

Pour plus d'information : écrire à info@editionscienceetbiencommun.org

Évaluation des interventions de santé mondiale

Méthodes avancées

Sous la direction de **Valéry Ridde** et **Christian Dagenais**

Une couverture universelle des soins de santé en 2030 pour tous les êtres humains, du Nord au Sud? Réaliser cet objectif de développement durable aussi ambitieux que nécessaire exigera une exceptionnelle volonté politique, mais aussi de solides données probantes sur les moyens d'y arriver, notamment sur les interventions de santé mondiale les plus efficaces. Savoir les évaluer est donc un enjeu majeur. On ne peut plus se contenter de mesurer leur efficacité : il nous faut comprendre pourquoi elles l'ont été (ou pas), comment et dans quelles conditions.

Cet ouvrage collectif réunissant 27 auteurs et 12 autrices de différents pays et de disciplines variées a pour but de présenter de manière claire et accessible, en français, un florilège d'approches et de méthodes avancées en évaluation d'interventions : quantitatives, qualitatives, mixtes, permettant d'étudier l'évaluabilité, la pérennité, les processus, la fidélité, l'efficience, l'équité et l'efficacité d'interventions complexes. Chaque méthode est présentée dans un chapitre à travers un cas réel pour faciliter la transmission de ces savoirs précieux.

esbc éditions science et bien commun

Un accès libre et universel, par le biais du numérique, à des livres scientifiques et documentaires publiés par des auteurs et des autrices de pays des Suds et du Nord, dans la perspective de la justice cognitive et de l'écologie des savoirs. Acheter un livre imprimé, c'est nous soutenir!

Editionscienceetbiencommun.org

ISBN au Québec : 978-2-924661-58-1
ISBN en France : 978-2-7099-2766-6

Design de couverture : Kate McDonnell, photographie de Christian Dagenais
Imprimé à Québec et Paris, 2019

ÉQUIPE RENARD RECHERCHE SUR LE TRANSFERT DE CONNAISSANCES

Institut de Recherche pour le Développement FRANCE

CC BY SA

Ce livre est aussi disponible en libre accès sous licence Creative Commons CC-BY-SA 4.0

9 782924 661581