



ELSEVIER
MASSON

Disponible en ligne sur
SciVerse ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com

Revue d'Épidémiologie
et de Santé Publique
Epidemiology and Public Health

Revue d'Épidémiologie et de Santé Publique xxx (2013) xxx-xxx

3^e Colloque thématique de l'Adelf
Épidémiologie sociale et inégalités de santé
Toulouse, 15 et 16 mai 2012

Pragmatisme et réalisme pour l'évaluation des interventions de santé publique

Pragmatism and realism for public health intervention evaluation

V. Ridde ^{a,*}, S. Haddad ^{a,b}

^a Centre de recherche du centre hospitalier de l'université de Montréal (CRCHUM), 1058 Saint-Denis, Montréal, QC H2L 4M1, Canada

^b Département de médecine sociale et préventive, faculté de médecine, université de Montréal, 7101, avenue du Parc, Montréal, QC H3N 1 × 7, Canada

Abstract

Forty years ago, Schwartz and Lellouch invented pragmatic clinical trials. Their proposal has not yet been fully espoused. This appears to be the case today also in the domain of public health interventions evaluation, where some still insist on the superiority of experimental methods. Yet evaluations of complex public health interventions are fraught with pitfalls for researchers. Most such interventions take place in natural experimental contexts, where they have no control over the context or the factors that modify implementation and influence the effects. Experimental approaches are, in these cases, not very appropriate, and yet decision makers want to be able to take decisions to improve them. This article presents our experience over the past 5 years with evaluative research in two public health interventions. We wish to show how we conduct evaluations in practice using a pragmatic approach. The article is focused on elements that have not, to date, received much attention in the francophone literature: the evaluability assessment and intervention logic, research strategies reinforced particularly by mixed methods and time series, and the analysis of implementation fidelity and mechanisms that foster effectiveness. Because the pragmatic approach to evaluative research stresses the need for good understanding of context and uses reinforced methodological strategies, it allows for rigorous responses to evaluation questions raised by those implementing complex public health interventions. Thus, experimental approaches are not necessarily required to analyze the effectiveness of interventions.

© 2013 Elsevier Masson SAS. All rights reserved.

Keywords: Evaluation; Complex interventions; Reinforced research strategies; Implementation; Mechanisms; Experiments; Pragmatism

Résumé

Il y a 40 ans, Schwartz et Lellouch ont inventé les essais cliniques pragmatiques. Cette proposition n'a pas encore été complètement prise en compte. Il semble que cela soit aujourd'hui aussi le cas dans le domaine de l'évaluation des interventions de santé publique où certains croient encore à la supériorité des méthodes expérimentales. Pourtant, la réalisation d'évaluation d'interventions complexes en santé publique est semée d'embûches pour les chercheurs. La majorité de ces interventions se déroulent dans des contextes d'expérimentations naturelles où ils n'ont pas le contrôle du contexte et des facteurs modifiant la mise en œuvre et influençant les effets. Les approches expérimentales sont alors peu pertinentes, alors que les responsables souhaitent prendre des décisions pour les améliorer. Cet article présente notre expérience des cinq dernières années de recherches évaluatives de deux interventions de santé publique. Nous voulons montrer comment, dans la pratique et de manière pragmatique, nous réalisons des évaluations. L'article est centré sur des éléments encore peu abordés dans les écrits francophones : la phase préévaluative et la logique d'intervention, les plans de recherche diversifiés par les méthodes mixtes et les séries chronologiques, l'analyse de la fidélité de l'implantation et des mécanismes propices à l'efficacité. Le recours à une approche pragmatique de recherche évaluative, en insistant sur la compréhension du contexte et en usant de stratégies méthodologiques renforcées, permet de répondre avec rigueur aux questions d'évaluation que se posent les responsables des interventions complexes en santé publique. L'analyse de l'efficacité des interventions ne doit donc pas nécessairement avoir recours à des approches expérimentales.

© 2013 Elsevier Masson SAS. Tous droits réservés.

Mots clés : Évaluation ; Interventions complexe ; Stratégie de recherche renforcée ; Implantation ; Mécanisme ; Expérimentations ; Pragmatisme

* Auteur correspondant.

Adresse e-mail : valery.ridde@umontreal.ca (V. Ridde).

0398-7620/\$ – see front matter © 2013 Elsevier Masson SAS. Tous droits réservés.

<http://dx.doi.org/10.1016/j.respe.2013.03.037>

Pour citer cet article : Ridde V, Haddad S. Pragmatisme et réalisme pour l'évaluation des interventions de santé publique. Rev Epidemiol Sante Publique (2013), <http://dx.doi.org/10.1016/j.respe.2013.03.037>

1. Introduction

Lorsqu'il s'agit d'évaluer des interventions de santé publique, le chercheur fait face à des objets complexes, car impliquant de multiples actions enchevêtrées, agissant dans des relations la plupart du temps non linéaires, organisées dans un contexte spécifique et comprises comme des systèmes ouverts dont le chercheur qui les observe fait lui-même partie [1]. Ainsi, la recherche évaluative, celle qui organise un processus systématique de collecte et d'analyse de données concernant une intervention dans le but de prendre des décisions [2], ne peut s'accommoder des représentations caricaturales et souvent polarisées – ce que Latour nommait les « divisions hautaines » [3] – entre méthodes quantitatives ou qualitatives, analyses déductives ou inductives ou encore chercheurs positivistes ou constructivistes. Dans ce type de pratique de recherche, les chercheurs en évaluation doivent souvent adopter une posture pragmatique pour répondre à leurs questions de recherche et fournir des recommandations aux décideurs [4].

2. Le pragmatisme en recherche : un concept ancien toujours en émergence

Récemment, les tenants des approches mixtes ont remis en débat le pragmatisme des années 1980 aux États-Unis [5,6]. Pour ceux-là, l'important concerne les conséquences de la recherche et de l'intervention étudiée, les questions de recherche guident les méthodes et non l'inverse. Ils peuvent avoir recours tant aux démarches inductives que déductives, tant à des données qualitatives que quantitatives. Leurs recherches sont surtout orientées par l'étude des pratiques et leur fonctionnement dans leur contexte. On reconnaît ici la perspective philosophique de Charles Peirce qui affirmait que l'importance d'une idée/action se juge à l'aune de la différence qu'elle peut provoquer dans le quotidien des populations [7]. Mais pour les épidémiologistes, notamment en France, ce retour du débat concernant le pragmatisme n'est pas vraiment nouveau. En effet, dans un article célèbre de la fin des années 1960 où les recherches expérimentales peinaient à produire des résultats utilisables dans d'autres contextes que ceux de leurs essais, Schwartz et Lellouch ont suggéré le développement d'expérimentations pragmatiques [8]. Selon eux, les stratégies de recherche mises en place lors « d'expériences (essais) interventionnelles robustes » [9] peuvent être organisées selon deux idéaux types (qui se situe dans un continuum et non en dichotomie¹). Le premier est qualifié d'expérimentations explicatives car celles-ci visent à évaluer l'efficacité théorique dans un contexte contrôlé et dans le but d'informer. Le second est qualifié d'expérimentations pragmatiques car elles sont mises en œuvre dans des contextes naturels, aussi proches que possible de la réalité, dans le but d'évaluer l'efficacité pratique mesurée en situation réelle pour soutenir la prise de décision [7,9]. En effet, « les décideurs efficaces valorisent le

¹ Des travaux subséquents proposent une liste de critères distinguant ces deux approches [8].

pragmatisme » [7] et il est important de leur apporter des réponses à partir d'interventions dont les résultats peuvent être transférés à d'autres contextes [10]. Ainsi, les questionnements contemporains des chercheurs en évaluation font-ils écho à leurs collègues réalisant des essais cliniques randomisés en recherche clinique. Mais il semble que l'appel à plus de pragmatisme réclamé il y a 40 ans n'ait trouvé qu'un écho limité en épidémiologie et en recherche clinique. En 2011, une revue spécialisée sur les essais cliniques a consacré plusieurs textes sur le sujet et soutenu que « l'expérimentation pragmatique est un concept émergent » [9]. Il nous semble que dans le domaine de l'évaluation, nous soyons également encore à l'émergence du concept et de son utilisation. Aussi, dans cet article, nous souhaitons nous éloigner des considérations théoriques et épistémologiques sur le pragmatisme scientifique abordés ailleurs [5,6], pour nous concentrer sur la présentation de notre expérience des cinq dernières années. Notre démonstration s'appuiera sur deux interventions qui se déroulent au Burkina Faso, l'une porte sur l'accès au système de santé et l'autre sur la lutte contre le paludisme. Nous avons choisi ces exemples car ils proviennent de notre expérience, mais les idées abordées dans cet article ne sont pas spécifiques à l'évaluation en Afrique. Faute de place, nous n'aborderons pas les résultats de nos travaux mais centrerons nos propos sur des éléments méthodologiques d'ordre pratique qui nous semblent avoir été peu abordés dans les écrits francophones.

3. Phase préévaluative : logique d'intervention et questions d'évaluation

Lorsque nous réalisons des évaluations, nous cherchons à répondre à des questions qui intéressent ceux qui vont potentiellement en utiliser les résultats pour améliorer les interventions, mais aussi contribuer à l'avancement des connaissances. Ce qui guide nos choix méthodologiques sont d'abord les questions d'évaluation et ensuite, évidemment, dans le contexte de rareté de ressources au sein duquel nous œuvrons, les aspects pratiques (la durée et le budget prévus pour la recherche ainsi que la prise de décision prévue avec ses résultats) [5,11]. Il faut donc commencer par définir des questions d'évaluation. C'est ici que la phase préévaluative a toute son importance.

La phase préévaluative est comprise comme l'étude de la faisabilité de la recherche évaluative. Elle est la phase qui permet aux chercheurs de préparer l'évaluation. À la fin des années 1970, Wholey [12] a proposé que quatre conditions soient étudiées lors de cette démarche préalable à l'évaluation :

- la logique d'intervention est claire, connue et partagée de tous ;
- les relations causales et la logique de l'intervention sont plausibles, autrement dit la théorie du changement a du sens et les activités planifiées devraient permettre d'atteindre les effets attendus ;
- les données nécessaires pour réaliser l'évaluation sont disponibles ou accessibles ;

- l'utilité de réaliser l'évaluation est reconnue par les utilisateurs potentiels des résultats.

Pour les chercheurs, ce dernier point est souvent peu pris en compte alors que l'utilisation des résultats de recherche est devenue cruciale dans une époque de politiques fondées sur des preuves [13]. La phase préévaluative est aujourd'hui une pratique courante [14] faisant partie des compétences essentielles des évaluateurs [15]. Mais cette pratique s'adapte essentiellement au contexte de sa mise en œuvre. Sa durée et sa profondeur varient en fonction des situations. Nos deux exemples empiriques, en nous centrant sur la logique d'intervention, le montrent parfaitement.

Dans le premier exemple, l'intervention consiste à supprimer la barrière financière à l'accès aux soins au Burkina Faso. Une ONG internationale financée par l'union européenne (ECHO) prend en charge la part financière qui est habituellement réclamée aux patients lorsqu'ils utilisent les services de santé. Depuis septembre 2008, les femmes enceintes, les enfants âgés de moins de cinq ans et les indigents ont un accès gratuit aux soins dans les formations sanitaires publiques de deux districts (Dori et Sebba) de la région sahélienne. Partant de l'expérience de notre collaboration antérieure d'une évaluation d'un programme similaire au Niger en 2007 [16], l'ONG a souhaité que nous l'accompagnions dans cette recherche évaluative au Burkina Faso. Si la phase préévaluative de l'été 2008 a été courte et centrée sur une description rapide de la logique d'intervention et des principales questions de recherche, il faut reconnaître que des retours permanents depuis 2008 ont été réalisés sur ces deux éléments. Autrement dit, la phase préévaluative n'a pas été fixée une fois pour toutes. Au fur et à mesure des années, la logique de l'intervention a été affinée, reprise, ajustée et les questions d'évaluation (voir infra) adaptées à l'évolution des besoins et des budgets disponibles. Le modèle logique de l'intervention (Fig. 1) est le fruit de ce processus itératif qui a permis aujourd'hui d'en arriver à une description très fine des processus et des effets attendus de l'intervention. Les liens étroits entre les chercheurs et les intervenants ainsi que la durée exceptionnelle du partenariat sur cinq années ont permis une telle description détaillée de la logique de l'intervention.

Le second exemple est le programme national de lutte contre le paludisme au Burkina Faso. Sans avoir réalisé une évaluation d'un projet pilote effectué dans trois districts du pays, les autorités ont obtenu en 2010 du Fond Mondial contre le paludisme, la tuberculose et le sida un budget de 63 millions d'Euros lui permettant d'étendre à l'échelle du pays une intervention de lutte contre le paludisme fondée sur deux stratégies. La première est une distribution massive (environ un ménage sur deux) et gratuite de moustiquaires imprégnées de longue durée d'action (MILDA). La seconde est la prise en charge communautaire à domicile (PECADO) des cas de paludisme simple à l'aide d'une combinaison thérapeutique à base d'artémisinine (ACT) vendu par les agents de santé communautaire dans les villages à un prix subventionné. Aucune évaluation nationale n'a été prévue. Nous avons donc décidé avec des chercheurs de l'Institut de recherche en science

de la santé (IRSS-CNRST) du Burkina Faso de réaliser une évaluation dans deux districts du pays (Kaya et Zorgho). La phase préévaluative a été plus laborieuse que dans le premier exemple, notamment parce que l'idée de mener une telle évaluation ne provenait pas des responsables du programme mais des chercheurs, contrairement à l'intervention précédente. Or, les démarches évaluatives sont encore trop souvent comprises comme des processus de sanctions. L'accès à l'information concernant les détails de l'intervention a donc été plus difficile. Fort de nos réseaux de connaissances locales, de nos multiples rencontres avec les parties prenantes aux niveaux central et local pour montrer que nous ne cherchions pas à porter de jugement sur des personnes ou des organisations en particulier, de nos visites de terrain dans certains districts et de quelques ateliers avec les responsables locaux de l'intervention, il a été possible de décrire la logique de l'intervention (Fig. 2).

De fait, la profondeur de cette phase préévaluative mais aussi la moindre complexité de l'intervention explique la présence d'un peu moins de détails sur la Fig. 2. Mais les détails ne sont pas toujours utiles et ce qui est surtout important est que la description de la logique d'intervention, la théorie du changement, permette :

- de comprendre les liens logiques entre les ressources, les activités et les effets attendus en théorie (d'où le terme « théorie d'intervention ») ;
- de sélectionner les questions d'évaluation dont les réponses permettront de vérifier la logique dans sa pratique, c'est-à-dire empirique ;
- de distinguer les questions relevant de différents types d'évaluation, par exemple des processus, des effets ou des impacts.

Ce dernier point est essentiel car on ne pourra affirmer qu'une intervention est efficace que si elle a produit des changements pour les bénéficiaires attendus, autrement dit, une réduction des dépenses de santé ou de la morbidité palustre par exemple. Le fait que les intervenants aient parfaitement distribué les moustiquaires ne relève pas de l'efficacité mais d'une analyse de l'implantation de l'intervention. Si ces distinctions ont l'air simple, elles ne sont pas toujours évidentes sur le terrain avec des intervenants pas toujours habitués aux démarches évaluatives. La représentation schématique de la logique d'intervention est un outil pédagogique souvent utile pour expliquer ces nuances.

Ainsi, c'est à partir des Fig. 1 et 2 que nous avons été en mesure de décider des questions de recherche évaluative. Encore une fois, le pragmatisme a guidé le choix des questions plutôt qu'une école de pensée. Pour évaluer des interventions de santé publique, nous adoptons ainsi une démarche tant pragmatique que dynamique, ce que Bourdieu [17] appelait de ses vœux il y a bien longtemps « la palette des méthodes utilisées [est] en adéquation avec le problème traité et [fait] l'objet d'une réflexion dans le mouvement même où on la déploie pour résoudre une question particulière ». Par exemple, certaines questions correspondent plus à une approche

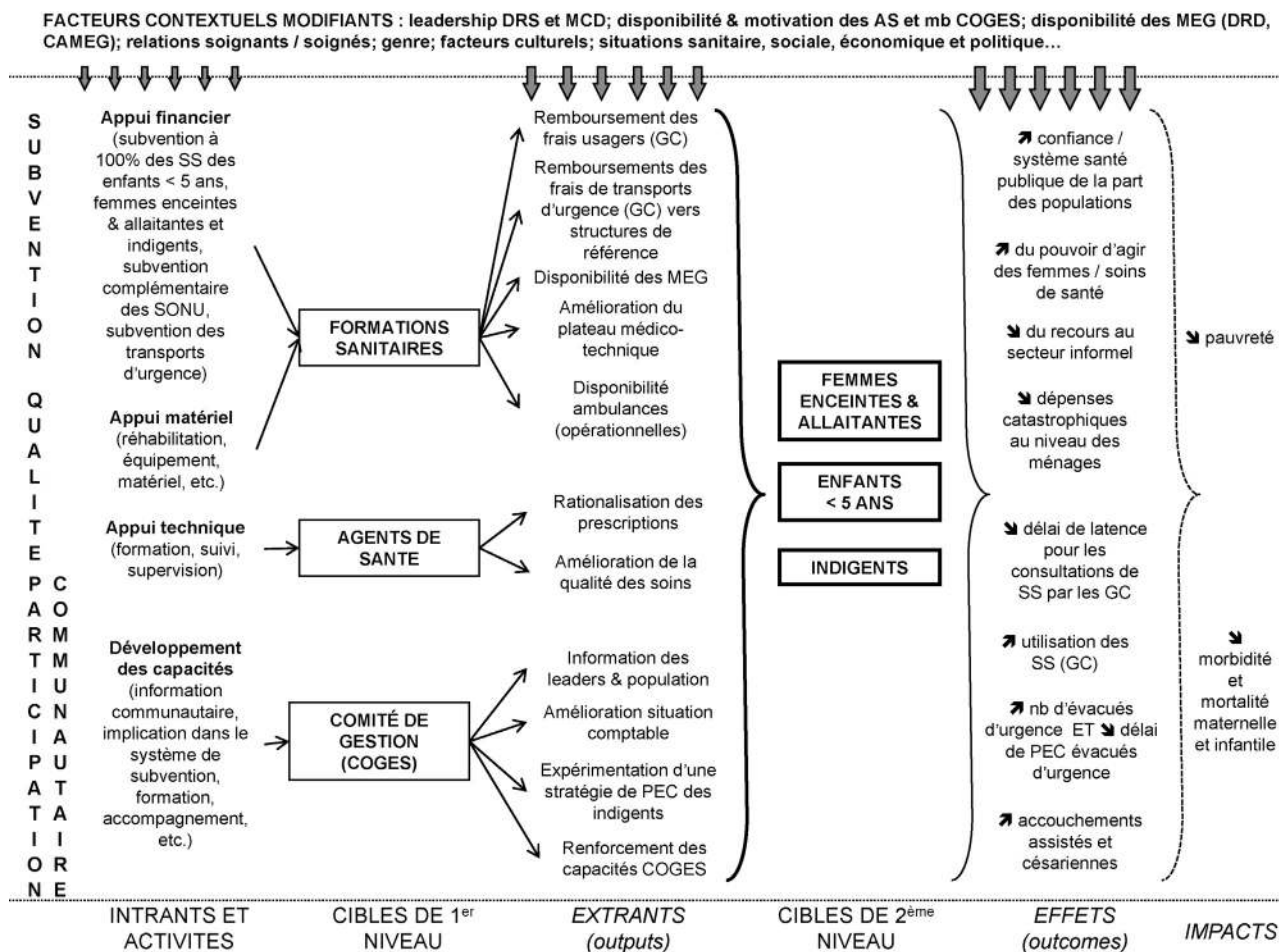


Fig. 1. Logique de l'intervention « Gratuité des soins ».

hypothético-déductive, notamment lorsqu'il s'agit d'évaluer l'efficacité des interventions, d'autres à une perspective holistico-inductive lorsque l'on cherche à comprendre les effets selon le point de vue des populations ou le processus de mise en œuvre des activités. Le **Tableau 1** présente une liste non exhaustive de questions d'évaluation concernant les deux interventions en fonction des types d'évaluation habituellement utilisées [2].

Le choix des questions d'évaluation est guidé par la logique d'intervention mais aussi par des considérations pragmatiques et politiques, c'est-à-dire essentiellement liées au budget et à la capacité de disposer de données pour répondre aux questions mais aussi aux besoins des utilisateurs des résultats [11]. Par exemple, le choix du district de Kaya pour évaluer la lutte contre le paludisme s'explique notamment car il existe sur ce site un système de surveillance démographique réalisant des autopsies verbales des décès d'enfants, seul moyen pour disposer de données de mortalité, ce qui n'est pas possible à Dori et Sebba. Notons que nous ne nous intéressons pas uniquement à l'efficacité des interventions mais aussi à leur équité, autrement dit à la distribution de leurs retombées entre les sous-groupes de la population.

4. Programme combinant plusieurs recherches

En ce qui concerne les deux interventions, le programme de recherches évaluatives a eu recours à un type complexe de recherche en méthodes mixtes que certains auteurs ont qualifié d'itératif car il s'agit de « combiner des méthodes qualitatives et quantitatives selon un processus dynamique, changeant ou évolutif tout au long du processus du programme de recherche » [18]. Selon ces auteurs [18], pour qu'un programme puisse ainsi être qualifié, il doit disposer des caractéristiques suivantes :

- un processus de recherche itératif et cyclique ;
- des décisions de départ qui influencent les suivantes ;
- une cohérence au sein des différentes étapes du programme de recherche ;
- des partenaires impliqués dans le processus de recherche ;
- une collaboration entre les différents chercheurs ;
- l'utilisation des perspectives qualitatives et quantitatives ;
- la prise en compte de plusieurs paradigmes.

En effet, c'est au fur et à mesure de ce programme de recherche que nous avons adapté les méthodes, que certaines réponses ont permis de se poser de nouvelles questions

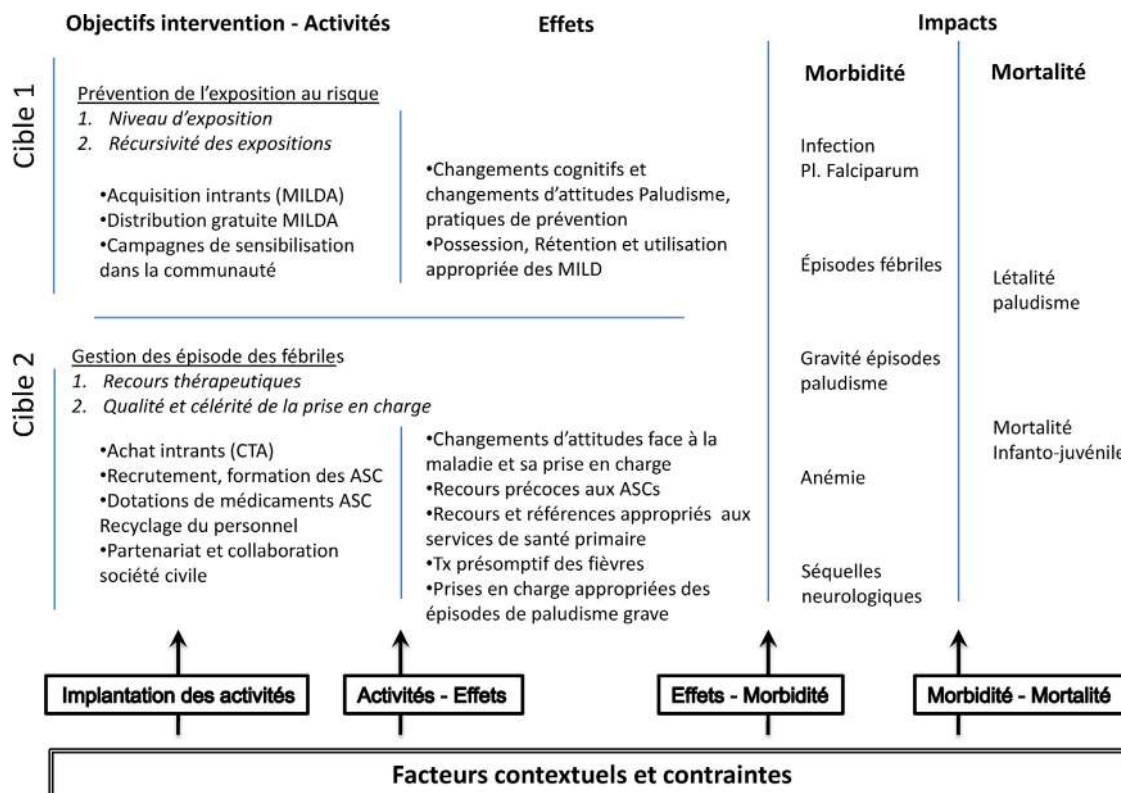


Fig. 2. Logique de l'intervention « Programme Palu ».

impliquant l'organisation, soit de nouvelles études, soit de nouvelles collectes de données au sein d'une même étude (voir infra). La faisabilité d'un tel processus itératif s'explique par l'ouverture d'esprit des responsables des interventions, par des budgets flexibles pouvant s'adapter au fur et à mesure des besoins exprimés par les parties prenantes mais aussi par la capacité de mobiliser une équipe de chercheurs ayant des

compétences interdisciplinaires (économie de la santé, sociologie, médecine, géographie, épidémiologie, etc.).

Une fois ces considérations préévaluatives présentées, il est évidemment impossible dans ces quelques pages de rendre compte de l'ensemble des détails de nos choix méthodologiques pour répondre aux questions d'évaluation avec pragmatisme. Aussi, nous allons essentiellement centrer les prochaines

Tableau 1
Exemples de questions de recherche concernant les deux interventions.

	La gratuité des soins . . .	La lutte contre le paludisme . . .
Impacts		réduit-elle la mortalité infantile due au paludisme ?
Effets	réduit-elle l'incidence des dépenses de santé dans les ménages et le risque d'appauvrissement dû à la maladie ? accroît-elle l'utilisation des formations sanitaires comme premier recours ? profite-t-elle à toutes les couches de la société (pauvres ou éloignés des centres) ? renforce-t-elle le pouvoir d'agir des femmes et des comités de gestion ? améliore-t-elle les capacités financières des comités de gestion ? a-t-elle bénéficié aux plus pauvres ?	réduit-elle la morbidité des enfants âgés de moins de 5 ans ? a-t-elle des impacts différentiels favorisant les enfants démunis ? a-t-elle augmenté la fréquentation des centres de santé primaires ?
Processus	a-t-elle été mise en œuvre selon un processus apprécié par les parties prenantes ? provoque-t-elle une charge de travail insurmontable pour le personnel de santé ? a-t-elle réussi à s'organiser en maintenant la qualité des prescriptions médicales et des soins obstétricaux ? les centres de santé sont-ils remboursés à temps par l'ONG ? pour les indigents a-t-elle provoqué une stigmatisation ?	a-t-elle été appréciée par les parties prenantes ? est-elle différemment mise en œuvre selon le contexte local ? est-elle implantée fidèlement au regard des activités et des échéanciers prévus ? a-t-elle intégré les agents de santé communautaire pour qu'ils soient efficaces ?
Pertinence	est-elle en phase avec les valeurs sociales en ciblant les femmes, les enfants et les indigents ?	

sections sur deux éléments relativement nouveaux dans les écrits francophones en recherche évaluative, soit le recours à des stratégies de recherche renforcées, l'analyse de la fidélité de l'implantation et la prise en compte du contexte.

5. Des plans de recherche diversifiés

Les plans de recherche diversifiés (ou renforcés) sont surtout utiles pour composer avec une double contingence de l'évaluation de l'efficacité des interventions :

- des sources d'hétérogénéité multiples liées au caractère « naturel » de l'expérimentation ;
- une demande pour des résultats allant au-delà de la simple mesure de l'efficacité.

Leur ambition est d'améliorer la stratégie de recherche afin de réunir un faisceau d'arguments permettant de soutenir les inférences causales et minimiser les menaces à la validité interne. Ces stratégies peuvent être opportunément combinées et relèvent principalement de trois logiques [19–21]. La première logique consiste à exploiter le potentiel de convergence d'observations ou de prédictions dérivées non pas d'une, mais de plusieurs sous-études, méthodes ou angles d'observation. Une combinaison judicieuse de plusieurs composantes, même imparfaites, est susceptible de réduire considérablement les inconvénients des plans quasi expérimentaux [20]. La seconde logique consiste à améliorer la validité interne d'un plan quasi expérimental n'ayant pas de groupe témoin grâce à l'extension des fenêtres d'observation, l'exploitation des sources d'hétérogénéité naturelles pour créer des « counterfactual » plus robustes (terme que l'on utilisera ici, faute de mieux²) ; et l'inclusion de variables dépendantes non équivalentes. Nous ne ferons que citer ici, sans la développer, la troisième logique qui relève de l'analyse des données par :

- un contrôle analytique des biais de sélection (appariements par scores de propension, modèles de type Heckitt, variables instrumentales, différence de différences) ;
- des modélisations adaptées à la nature hiérarchique des données (modèles multiniveaux à coefficients aléatoires) ;
- des analyses de sensibilité complexes [22,23].

5.1. L'utilisation de multiples critères de résultats

Le but est de sélectionner un large ensemble d'indicateurs plutôt que de fonder l'analyse de l'efficacité sur une mesure de résultat unique faisant fonction de mesure de résultats principale. L'objectif est d'éviter de ne prendre que des indicateurs d'impacts mais d'explorer aussi les résultats intermédiaires. La démarche permet d'examiner l'ensemble des mécanismes putatifs de production des effets. Cette

² Il n'existe aucun équivalent en langue française du terme « counterfactual », désignant ce qu'aurait été la condition (non mesurable) des personnes exposées s'ils ne l'avaient pas été.

démarche de sélection des mesures de conséquences est conditionnée par une spécification préalable de la théorie du changement (voir supra). Prenons l'exemple de la Fig. 2. Il s'agit de disposer d'indicateurs de résultats illustrant le mécanisme ciblant la prévention de l'exposition au risque et celui couvrant la gestion des épisodes fébriles (logique verticale). Il faut ensuite, pour chaque mécanisme, disposer d'indicateurs couvrant l'implantation, les extrants, les effets proximaux (attitudes, comportements, pratiques de soins), la morbidité et enfin, la mortalité (logique horizontale). La reconstruction de la séquence d'effets permet par la suite de valider la mesure des effets (les changements observés doivent être cohérents) et pour mieux comprendre et expliquer, de mettre en lien les changements observés avec les réalisations.

5.2. Le recours à plusieurs composantes

La combinaison de plusieurs plans quasi expérimentaux imparfaits peut réduire les menaces sur la validité interne si la mise en commun de leurs résultats montre une réplification littérale (les résultats sont comparables) ou théorique (les observations confirment l'existence de différences anticipées avant l'analyse). Ainsi, à Dori (Fig. 1), deux stratégies sont utilisées pour évaluer les effets de l'intervention sur la fréquentation des centres de santé par les enfants. La première repose sur un plan de type avant-après non aléatoire, sans groupe témoin. L'enquête de panel (échantillon aléatoire de ménages) permet de rendre compte de la progression des recours aux soins dans les deux sites d'intervention. Mais ce plan quasi expérimental est très sensible aux biais historiques (histoire, maturation, attrition, réactions compensatoires, régression vers la moyenne) et ne permet pas de fournir des preuves suffisamment robustes de l'efficacité de l'intervention. Une analyse longitudinale dans 12 centres de santé (six dans un site d'intervention et six dans un site témoin) a aussi été réalisée en se fondant sur des séries longitudinales interrompues (voir infra). L'échantillon des centres de santé n'est pas aléatoire et il y a un risque de biais de sélection. Comme pour toute série longitudinale, des biais d'instrumentation sont aussi possibles, et on ne peut exclure une modélisation imparfaite de l'évolution de confondants potentiels. Mais finalement, les deux méthodes fournissent des estimés très proches de l'effet de l'intervention. Cette mise en commun des résultats ne fournit sans doute pas de garanties définitives sur la validité des résultats, elle contribue toutefois à crédibiliser les inférences causales.

5.3. Le recours à des méthodes mixtes

Associer des approches qualitatives et quantitatives est une logique de renforcement des stratégies de recherche de plus en plus courante et s'appuie désormais sur des typologies plus formelles [6]. On distingue habituellement quatre approches en méthodes mixtes (Tableau 2) qui s'appliquent pour chaque étude plutôt qu'à des programmes de recherche complexe [24].

Pour illustrer notre utilisation pragmatique de ces différents protocoles, nous présentons dans le Tableau 3 la série d'études employée lors de l'évaluation de l'intervention de gratuité des

Tableau 2
Les quatre principales approches en méthodes mixtes.

Stratégie	Temporalité	Prépondérance	Analyse	Notation
Triangulation	Concomitant	Égale	Fusionner les données lors de l'analyse	QUAN + QUAL
Imbriquée	Concomitant ou séquentielle	Inégale	Imbriquer un type de donnée dans l'autre	QUAN (qual) ou QUAL (quan)
Explicative	Séquentielle, quantitatif d'abord	QUAN	Relier les deux types de données	QUAN → qual
Exploratoire	Séquentielle, qualitatif d'abord	QUAL	Relier les deux types de données	QUAL → quan

Adapté de Pluye, 2012 [24].

soins au Sahel (Fig. 1) en mentionnant le type d'étude au regard de l'instrument proposé par Pluye et al. [25] et l'importance relative des données qualitatives (QUAL) et quantitatives (QUANT). On comprendra que certaines études ont été purement QUAL ou QUANT tandis que d'autres ont été de nature mixte avec une prépondérance différente des deux types de données.

Prenons un exemple pour illustrer notre démarche de méthodes mixtes. L'idée de l'étude 5 sur la qualité des prescriptions médicales est née de l'inquiétude (qualitative) parfois exprimée par les agents de santé que la gratuité allait engendrer une mauvaise qualité des soins. Pour vérifier cette hypothèse, nous avons donc entrepris une étude quantitative (voir plus bas) sur la qualité des prescriptions médicales au regard des normes nationales et de l'Organisation mondiale de la santé (OMS) [26]. Parallèlement, pour mieux comprendre le point de vue des prescripteurs, nous avons en même temps effectué des entretiens approfondis avec un échantillon d'agents de santé qui pouvaient ainsi exprimer leur perception des changements apportés par l'intervention.

5.4. Explorer les sources naturelles d'hétérogénéité

En pratique, les sources d'hétérogénéité sont nombreuses et trouvent leur origine dans la variabilité des interventions, des

contextes dans lesquels elles sont déployées, et des réponses des populations à ces interventions. La question de l'hétérogénéité va donc bien au-delà de la problématique de l'implantation partielle du traitement [20], dont on essaie de tenir compte dans les essais randomisés pragmatiques. L'hétérogénéité n'est donc pas comme l'indique Deaton [27] « juste un problème technique » à résoudre, mais une dimension clé de la pratique d'évaluation. Plutôt que de la traiter comme une « nuisance » [28], il s'agit ici d'exploiter l'hétérogénéité naturelle pour créer des comparateurs appropriés. Une démarche courante consiste à procéder dans une logique de type dose-effets en vérifiant l'existence d'un gradient d'effets correspondant à des variations du degré d'implantation dans différents contextes. De la même manière, les résultats observés peuvent être mis en relation avec des variations dans la qualité des interventions, leur couverture ou la participation des populations. La démarche peut être étendue de manière à vérifier un réseau d'hypothèses associant processus-contexte-résultats. Cette approche est particulièrement utile car elle permet de proposer des « counterfactuals » plus nombreux et plus robustes en démultipliant les contrastes dans des situations où la singularité de l'intervention ou du contexte, limite la capacité de l'évaluateur à incorporer des sites témoins (ou encore, dans le cas de programmes couvrant l'intégralité d'un territoire).

Tableau 3
Liste des études entreprises et importance relative des données qualitatives et quantitatives.

	Thème des études	Type d'études	Fenêtre ou période	Données	Échantillon
1	Panel de ménages	Cohorte	2008, 2009, 2012	QT	2210 ménages
2	Utilisation dans les centres de santé primaires	M : explicatif	2004–2012	QUAN → qual	72 CSPTS
3	Contexte communautaire et centres de santé primaire	M : triangulation	2004–2009	QT, QL	28 CSPTS
4	Finances des comités de gestion	Étude cas-témoins	2008–2009	QT	28 CSPTS
5	Qualité des prescriptions médicales	M : triangulation	2008–2009	QT + QL	13 059 ordonnances et 14 prescripteurs
6	Ciblage des indigents	M : explicatif	2009–2010	QUAN → qual	450 et 140 personnes
7	Dépenses et perception de la qualité des accouchements	M : triangulation	2010	QT + QL	849 parturientes
8	Participation communautaire et pouvoir d'agir	Qualitative	2011	QL	140 personnes
9	Délais remboursements	Étude de cas	2009	QT, QL	10 CSPTS
10	Coûts des actes remboursés	Quantitative descriptive	2009	QT	28 CSPTS
11	Processus de mise en œuvre	M : triangulation	2009	QL + QT	436 personnes
12	Paiement au forfait	M : triangulation	2011–2012	QT + QL	6 CSPTS
13	Charge de travail du personnel soignant	M : triangulation	2011	QT + QL	8 CSPTS et 26 agents de santé
14	Passage à l'échelle et pérennisation	Qualitative	2012	QL	53 personnes

M : étude mixte ; QT : quantitative ; QL : qualitative ; QL_T : quali/quant minoritaire.

5.5. Inclusion de variables dépendantes non équivalentes

Cette démarche, formalisée par Shadish et Cook, consiste à procéder à une réplification théorique sur deux variables dépendantes, l'une sensible à l'intervention et l'autre ne l'étant pas [22]. La variable non sensible à l'intervention a cela de particulier qu'elle doit être exposée à des biais de validité interne identiques à ceux que rencontre la première. Les deux variables étant exposées aux mêmes facteurs de confusion, les éventuelles différences observées à l'intervention sont plus facilement attribuables à l'intervention. Nous présentons un exemple de ce type de démarche dans la section suivante.

5.6. Étendre les fenêtres d'observation et utiliser des séries chronologiques interrompues

Les difficultés à trouver des comparateurs explique le recours fréquent à des plans d'évaluation de type avant-après sans groupe témoin (O1 X O2), dont on a vu qu'ils étaient très sensibles aux biais d'histoire (et au biais de sélection dans la mesure où l'assignation randomisée des sujets est l'exception) [29]. L'extension de la fenêtre d'observation par un double prétest permet de mieux qualifier la situation et les processus naturels dans la période précédant l'intervention. Une appréciation plus fine des menaces sur la validité interne attribuables à des processus de maturation, ou de régression vers la moyenne, conduit à une meilleure spécification du « counterfactual ». Des observations répétées après l'implantation fournissent de meilleurs estimés des effets attribuables à l'intervention tout en permettant une meilleure analyse des pertes de vue en cours d'étude [30]. Des intervalles réduits entre ces observations tendent à réduire les biais d'histoire (survenue d'événements externes). La validité interne va s'accroître à mesure que se multiplient les observations et s'étend la fenêtre d'observation.

Les séries chronologiques interrompues consistent en de multiples observations séparées par des intervalles habituellement équidistants et « interrompus » par une intervention. L'évaluation repose sur la comparaison des segments précédant et suivant l'implantation de l'intervention. Le premier segment rend compte de la progression et de l'évolution naturelle (tendance séculaire, cycles longs, saisonnalité) du phénomène d'intérêt. Sa modélisation sert donc principalement à établir un « counterfactual » robuste et à contrôler les biais de validité interne [31]. La modélisation du second segment sert à évaluer les effets de l'intervention. L'analyse s'attache à identifier une rupture entre les segments : un saut (effet d'« intercept ») ou une modification immédiate ou décalée de la tendance séculaire initiale (effet de pente). La modélisation repose sur des techniques statistiques tenant compte notamment, de l'auto-corrélation temporelle et de la structure hiérarchique des données.

Les séries interrompues nécessitent un nombre d'observations suffisant pour pouvoir établir les formes fonctionnelles des segments. Leurs estimés sont meilleurs lorsque l'implantation de l'intervention est franche et réalisée sur une courte période. Elles ne permettent pas d'éliminer toutes les menaces à la validité interne [22], mais elles constituent une alternative de

choix aux plans quasi expérimentaux de type pré-post, et sont, de ce fait, extrêmement utilisées en évaluation. Le « Cochrane Effective Practice and Organisation of Care Review Group (EPOC) » en reconnaît désormais la pertinence pour la réalisation d'études systématiques. La validité interne des séries interrompues peut être améliorée par l'analyse simultanée de séries observées dans différents sites d'intervention, l'adjonction de séries témoins, de variables dépendantes non équivalentes, ou des stratégies plus complexes telles que l'observation après retrait de l'intervention ou inversion des sites d'exposition [20,22].

5.7. Deux exemples de plan de recherche renforcé

La Fig. 3 rend compte des analyses réalisées dans la région de Dori pour estimer les effets à court et moyen terme de deux interventions visant à accroître l'accès à l'accouchement par du personnel qualifié dans les formations sanitaires. La première intervention consiste à subventionner le prix de l'accouchement dans les centres de santé (réduction d'environ 80 % du prix). Elle concerne les quatre districts de la région sanitaire. La seconde consiste à étendre la subvention de manière à ce que l'accouchement soit gratuit (Fig. 1). Elle ne concerne que deux districts (Sebba et Dori). Les données proviennent des services des statistiques d'information sanitaire dont on a vérifié la qualité. L'indicateur de résultat est le nombre d'accouchements mensuels réalisés un mois donné dans chacun des 80 centres de santé des quatre districts. La fenêtre d'observation couvre une période de 96 mois (janvier 2004–décembre 2011). L'analyse est réalisée par modélisation multiniveaux en regroupant les 80 séries temporelles. La modélisation est non linéaire pour tenir compte des effets de plafond. La Fig. 3a présente la moyenne d'accouchements par centre observée dans chacun des districts pendant la période d'observation. Les lignes verticales rendent compte de la date d'implantation des deux interventions. La Fig. 3b présente les résultats de la modélisation. Les courbes représentent les gains attribuables à chaque intervention et rendent compte des sauts (« intercepts ») et des modifications de tendance (pente). La Fig. 3c rend compte des gains attribuables au cumul des deux interventions.

L'exemple suivant porte aussi sur la région de Dori. Il illustre une démarche associant séries chronologiques interrompues, inclusion de variables dépendantes non équivalentes et analyse multicritères. L'évaluation porte sur l'effet présumé de la gratuité des soins aux enfants sur la qualité des prescriptions médicales (Fig. 1, étude 5 supra) [26]. Certains décideurs craignaient en effet, que la gratuité ne conduise les agents à prescrire davantage de médicaments plus coûteux et notamment des antibiotiques et des injectables. La première variable dépendante est le taux de prescription d'antibiotiques pour des enfants âgés de 0 à 5 ans (les seuls éligibles à l'intervention) se présentant avec un paludisme simple (ne requérant pas d'antibiothérapie). Si la qualité se détériorait, on observerait un accroissement du taux de prescription mensuel d'antibiotiques aux enfants éligibles dans le segment post-intervention. Une variable non équivalente a ensuite été ajoutée : le taux de prescription d'antibiotiques aux enfants âgés de cinq à neuf

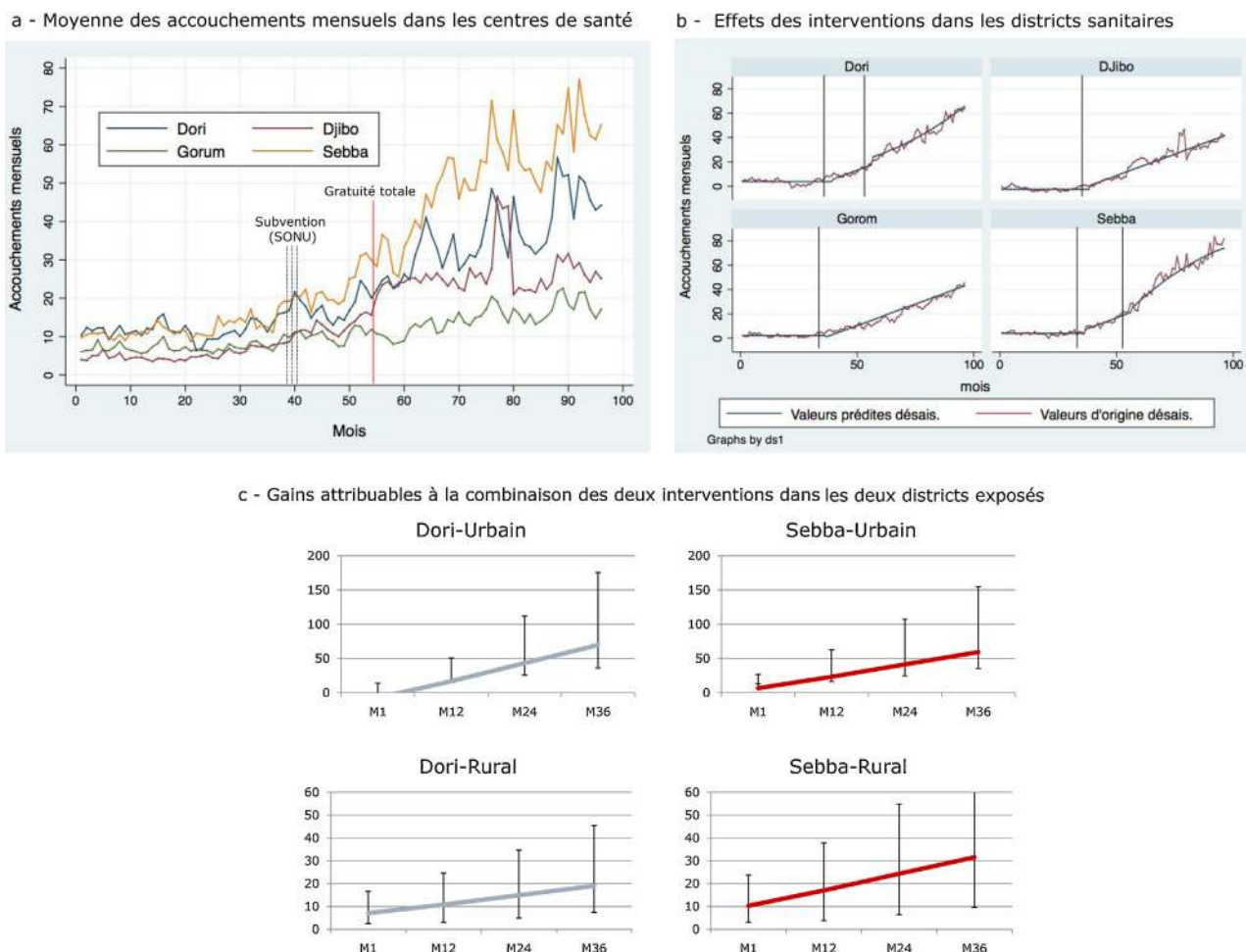


Fig. 3. Utilisation de séries chronologiques interrompues : a : moyenne des accouchements mensuels dans les centres de santé ; b : effets des interventions dans les districts sanitaires (données désaisonnalisées) ; c : gains attribuables à la combinaison des deux interventions dans les deux districts exposés après un à 36 mois d'intervention.

ans (non éligibles à la gratuité). Ces deux variables sont pour une large part sujettes à des confondants similaires car relevant des pratiques des mêmes agents dans les mêmes centres de santé et mesurées au même moment. Comme nous l'anticipions et contrairement aux craintes formulées, l'intervention n'a pas conduit à une modification mesurable des pratiques de prescription ciblant les enfants éligibles. La variable comparative a montré un profil d'évolution identique. Cette démarche a ensuite été reproduite sur deux autres indicateurs de qualité : les taux de prescriptions conformes et de prescription d'injectables. Les résultats des différentes formes de répliation testées ont fourni des résultats très congruents. La combinaison des trois démarches méthodologiques (extension de la fenêtre d'extension, indicateurs multiples de résultats et variables dépendantes non équivalentes) a contribué à démontrer que l'intervention n'avait vraisemblablement pas induit de détérioration dans la qualité des prescriptions.

6. L'analyse de l'implantation pour éviter l'erreur de type « trois »

Même si les questions d'implantation ne leur sont pas étrangères [31], les lecteurs sont certainement plus habitués à

entendre parler des erreurs de type un et deux que d'erreur de type trois. Pour les évaluateurs, notamment ceux qui s'intéressent à l'efficacité des interventions dans un contexte naturel [32], cette dernière est pourtant essentielle. Il ne faudrait pas en effet commettre l'erreur d'évaluer une intervention alors qu'elle n'a pas été mise en œuvre telle que prévu [33]. L'analyse d'implantation permet ainsi notamment d'éviter les erreurs dites de type trois.

Dans le contexte de l'intervention de lutte contre le paludisme, la mise en œuvre des deux volets vise ultimement à réduire l'incidence du paludisme et à sauver des vies. L'efficacité théorique de cette approche a été prouvée dans des contextes d'expérimentation. En revanche, l'efficacité en situation réelle n'est pas suffisamment connue. Or, si on se fie à l'histoire des programmes de lutte contre le paludisme au Burkina Faso, il est permis de s'attendre à des écarts sensibles dans l'implantation, [34]. Nous avons donc décidé de nous intéresser d'abord à la fidélité de l'implantation [35,36]. Il s'agit d'une étape préalable à l'évaluation du processus qui permet de mieux comprendre, lors d'une étude ultérieure, la dynamique du programme, le jeu des acteurs et les facteurs ayant contribué au degré d'implantation constaté [37]. L'évaluation a été effectuée au moyen d'une étude de cas

multiples avec plusieurs niveaux d'analyses imbriqués correspondant aux dimensions de la fidélité [38]. La phase préévaluative décrite plus haut a ici toute son importance car elle a permis de dresser la liste des activités prévues à partir des documents de planification du programme et de huit entretiens avec les principales parties prenantes de l'intervention. Elles ont validé cette liste qui concerne 12 activités pour le volet préventif (MILDA) et 35 pour le volet curatif (PECADO). Pour l'analyse du contenu réalisée avec un accord inter-juge, à partir de données empiriques (observations de terrain durant dix semaines, analyse documentaire, 48 entretiens individuels), chacune des activités réalisées dans 36 villages et deux villes a été évaluée en fonction de cinq modalités : implantée telle que prévue, modifiée, implantée ou modifiée, ajoutée, non implantée. Pour l'analyse de la couverture des activités, nous avons étudié si le nombre d'activités planifiées et le territoire concerné avaient été respectés. Enfin, pour l'analyse de l'échéancier, nous nous sommes concentrés sur les activités pour lesquelles des dates et délais avaient été préconisés.

Dans la liste des activités qui concerne la PECADO, le rôle des agents de santé communautaire (ASC) est essentiel car ce sont eux qui sont en première ligne pour soigner les enfants dans les villages dans le cas de paludisme simple. L'efficacité de cette stratégie, pourtant abandonnée depuis de nombreuses années au Burkina Faso faute de résultats probants et de conditions pour en assurer la pérennité [39], méritait une attention particulière. C'est ainsi que nous avons décidé de réaliser une évaluation qui s'inscrit dans une démarche d'approche Realist [40] où le contexte, les acteurs et les mécanismes sont des facteurs modérateurs importants de l'implantation et, in fine, de l'efficacité des interventions. L'approche a des méthodes d'évaluation fondées sur la théorie [41] peut être considérée comme une « mechanism-based approach » [42]. L'approche dispose souvent un caractère exploratoire et permet de mettre au jour, tester puis améliorer des théories dites de moyenne portée [43]. Ces théories concernent des mécanismes de changements qui sont liés aux raisonnements humains et concernent les choix des personnes ainsi que leurs aptitudes [41,44]. Cette notion de mécanisme mobilisé par les ASC organisant la PECADO est le cœur de notre étude. Car, si les programmes fonctionnent, c'est bien parce que les personnes en ont décidé ainsi et non pas simplement car des ressources sont mobilisées ou des activités sont mises en place. Il s'agit donc de mettre en lumière le mécanisme (M) actionné par l'intervention pour produire ses effets (E) dans un contexte donné (C), ce que l'on appelle la configuration « Contexte-Mécanisme-Effet » (CME). Puis, il s'agit de comparer les CME à la « théorie de l'intervention », autrement dit la logique de l'intervention mise au jour lors de la phrase préévaluative (voir supra). Rendre compte de la manière dont fonctionne la logique d'intervention (et les CME) dans un contexte spécifique avec ce cadre d'analyse permet de projeter les résultats de recherche dans différents contextes. Partant du postulat qu'il est envisageable d'observer différentes interventions déployant la même logique — une même famille de mécanismes — dans des contextes différents, l'approche suggère la possibilité de mettre au jour des régularités (*patterns*) — liées à des caractéristiques spécifiques du contexte — dans les effets

constatés. C'est en réalisant de telles observations empiriques (ou des synthèses) multiples dans des contextes variés qu'il est a priori possible de voir apparaître ces régularités.

Dans un premier temps, sur la base d'une recension des écrits d'interventions ayant utilisé une approche expérimentale d'évaluation comprise par certains comme l'étalon de la recherche causale permettant de maximiser la validité interne de l'étude [31], pour tester l'efficacité des ASC, les chercheurs ont mis au jour la présence de sept mécanismes a priori efficaces [45] :

- le sentiment d'appartenance et de compte à rendre au système de santé ;
- le sentiment d'appartenance et de compte à rendre à la communauté ;
- le sentiment de crédibilité et de légitimité du fait d'appartenir au système de santé ;
- l'anticipation d'être valorisé par le système de santé ;
- l'anticipation d'être valorisé par la communauté ;
- une perception de l'amélioration de son statut social et d'avoir d'un rôle social ;
- une assurance qu'il existe un système de rétro-action et de soutien à son activité.

Puisque ces mécanismes sont a priori efficaces, nous avons réalisé des entretiens auprès des ASC pour en percevoir et décrire la présence. Puis, pour renforcer la validité des résultats la collecte des données est étendue à d'autres personnes, les mères, les chefs de village et les agents de santé. Dans le cas de cette recherche, nous disposons d'une rare opportunité d'utiliser cette approche Realist dans sa totalité et surtout dans son originalité de mettre au jour les configurations CME. En effet, les enquêtes ménages vont nous permettre de disposer de mesure d'impacts objectives (morbidité, anémie, parasitémie, malnutrition) liées au paludisme. Car si cette approche méthodologique est prometteuse, il est exceptionnel de disposer d'un terrain empirique permettant d'avoir des données pour documenter les CME dans leur ensemble, il manque la plupart du temps surtout le E [40,46]. Ainsi, les analyses quantitatives devraient permettre de mettre au jour l'évolution des états de santé différents selon les aires de santé où interviennent les ASC des deux districts. Cette évolution pourra en partie être comprise comme des changements induits par les interventions au sein desquelles les ASC sont fortement impliqués. Évidemment, ces états de santé peuvent s'expliquer par la présence d'autres mécanismes que ceux liés aux ASC mais aussi par leur contexte de mobilisation. Des entretiens en profondeur seront donc effectués auprès des principales parties prenantes des interventions afin de documenter d'éventuelles récurrences empiriques de type C-M, M-E, C-E ou mieux, des CME pour tenter de disposer de théories de moyenne portée.

7. Conclusion

Notre expérience dans l'évaluation de ces deux interventions complexes de santé publique montre qu'il faut certainement s'éloigner du dogmatisme des approches paradigmatiques pour

se diriger vers plus de pragmatisme. L'appel de Schwartz et Lellouch des années 1960 pour que les chercheurs prennent des postures plus pragmatiques dans les expérimentations mérite certainement d'être de nouveau lancé et aussi entendu par la communauté des chercheurs et des décideurs en évaluation d'interventions de santé publique. L'approche réflexive entreprise dans cet article mérite d'être poursuivie afin de renforcer nos pratiques de recherche évaluative pour mieux répondre aux questions des décideurs en santé publique, ce qui est aussi une responsabilité scientifique.

Déclaration d'intérêts

Les auteurs déclarent ne pas avoir de conflits d'intérêts en relation avec cet article.

Remerciements

Les évaluations de ces deux interventions sont financées par les Instituts de recherche en santé du Canada (IRSC), le Fonds de recherche du Québec – Santé (FRQS), et European Community Humanitarian Office (ECHO) en collaboration avec l'ONG allemande HELP. Valéry Ridde est « Nouveau Chercheur » des IRSC. Merci à Ludovic Queuille, Thomas Druetz, Kadidiatou Kadi, Seni Kouanda et tous nos collègues pour leur collaboration dans l'évaluation des interventions dont il est question dans cet article réflexif dont nous prenons seuls la responsabilité.

Références

- [1] Blaise P, Marchal B, Lefèvre P, Kegels G. Au-delà des méthodes expérimentales, l'approche réaliste en évaluation. In: Potvin L, Moquet MJ, Jones C, editors. Réduire les inégalités sociales de santé. Collection « Santé en action ». Saint-Denis: INPES; 2010. p. 285–96.
- [2] Ridde V, Dagenais C, editors. *Approches et pratiques en évaluation de programme*. 2^e éd., Montréal: Presses de l'Université de Montréal; 2012.
- [3] Latour B. Culture et technique. Les « vues » de l'esprit. *Réseaux* 1987;5(27):79–96.
- [4] Patton MQ, LaBosière F. L'évaluation axée sur l'utilisation. In: Ridde V, Dagenais C, editors. *Approches et pratiques de l'évaluation de programme*. 2nd ed., Montréal: Presses de l'Université de Montréal; 2012. p. 145–60.
- [5] Robson C. *Real world research: a resource for users of social research methods in applied settings*, 3rd ed., Chichester (West Sussex): Wiley; 2011 [xxi, 586 p].
- [6] Tashakkori A, Teddlie C. *Sage handbook of mixed methods in social and behavioral research*, 2nd ed., Los Angeles: SAGE Publications; 2010 [xv, 893 p].
- [7] Maclure M. Explaining pragmatic trials to pragmatic policymakers. *J Clin Epidemiol* 2009;62:476–8.
- [8] Treweek S, Zwarenstein M. Making trials matter: pragmatic and explanatory trials and the problem of applicability. *Trials* 2009;10:37.
- [9] Patsopoulos NA. A pragmatic view on pragmatic trials. *Dialogues Clin Neurosci* 2011;13:217–24.
- [10] Cambon L, Minary L, Ridde V, Alla F. Transferability of interventions in health education: a review. *BMC Public Health* 2012;12(1):497.
- [11] Bamberger M, Rugh J, Mabry L. *Real world evaluation: working under budget, time, data and political constraints*. Thousand Oaks (CA): Sage Publications; 2006, 504 p.
- [12] Wholey JS, editor. *Evaluability assessment*. 2nd ed., San Francisco: Jossey-Bass; 2004.
- [13] Dagenais C, Robert É, editors. *Le transfert des connaissances dans le domaine social*. Montréal: Presses de l'Université de Montréal; 2012.
- [14] Trevisan MS. Evaluability assessment from 1986 to 2006. *Am J Eval* 2007;28(3):290–303.
- [15] Stevahn L, King JA, Ghere G, Minnema J. Establishing essential competencies for program evaluators. *Am J Eval* 2005;26(1):43–59.
- [16] D'Ostie-Racine L, Dagenais C, Ridde V. An evaluability assessment of a West Africa based Non-Governmental Organization's (NGO) progressive evaluation strategy. *Eval Program Plann* 2012;36(1):71–9.
- [17] Bourdieu P. *Science de la science et réflexivité*. Paris: Raisons d'agir; 2001, 238 p.
- [18] Nastasi BK, Hitchcock JH, Brown LM. An inclusive framework for conceptualizing mixed methods design typologies: moving toward fully integrated synergistic research models. In: Tashakkori ACT, editor. *Sage handbook of mixed methods in social and behavioral research*. Thousand Oaks (CA): Sage Publications; 2010. p. 305–38.
- [19] Angrist J, Pischke J. The credibility revolution in empirical economics: how better research design is taking the con out of econometrics. *J Econ Perspect* 2010;24(2):3–30.
- [20] Shadish WR, Cook TD. The renaissance of field experimentation in evaluating interventions. *Annu Rev Psychol* 2009;60:607–29.
- [21] Mercer S, DeVinney B. Study designs for effectiveness and translation research identifying trade-offs. *Am J Prev Med* 2007;33(2):139–54.
- [22] Shadish WR, Cook TD, Campbell DT. *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin; 2002 [xxi, 623 p].
- [23] Angrist J, Pischke J. *Mostly harmless econometrics: an empiricist's companion*. Princeton University Press; 2008.
- [24] Pluye P. Les méthodes mixtes pour l'évaluation des programmes. In: Ridde V, Dagenais C, editors. *Approches et pratiques en évaluation de programme (seconde édition)*. Montréal: Presses de l'Université de Montréal; 2012.
- [25] Pluye P, Robert E, Cargo M, Bartlett G, O'Cathain A, Griffiths F, et al. Proposal: a mixed methods appraisal tool for systematic mixed studies reviews. Available from: <http://mixedmethodsappraisaltoolpublic.pbworks.com/f/MMAT2011criteriaandtutorial2011-06-29.pdf>. Retrieved on [date] from <http://mixedmethodsappraisaltoolpublic.pbworks.com>. Archived by WebCite[®] at <http://www.webcitation.org/5tTRTC9yJ>; 2011 [cited 2011 Aug 22].
- [26] Atchessi N, Ridde V, Haddad S. Combining user fees exemption with training and supervision helps to maintain the quality of drug prescriptions in Burkina Faso. *Health Policy Plann* 2012 (In press).
- [27] Deaton A. Understanding the mechanisms of economic development. *J Econ Perspect* 2010;24(3–Summer 2010):3–16.
- [28] Davidoff F. Heterogeneity is not always noise: lessons from improvement. *JAMA* 2009;302(23):2580–6.
- [29] Campbell DT, Stanley JC. *Experimental and quasi-experimental designs for research*. Chicago (Illinois): Rand McNally; 1963, 84 p.
- [30] Cook TD, Campbell DT. *Quasi-experimentation: design and analysis for field settings*. Boston: Houghton Mifflin; 1979, 420 p.
- [31] Mercer SL, DeVinney BJ, Fine LJ, Green LW, Dougherty D. Study designs for effectiveness and translation research identifying trade-offs. *Am J Prev Med* 2007;33(2):139–54.
- [32] Petticrew M, Cummins S, Ferrell C, Findlay A, Higgins C, Hoy C, et al. Natural experiments: an underused tool for public health? *Public Health* 2005;119(9):751–7.
- [33] Dobson D, Cook TD. Avoiding type III error in program evaluation: results from a field experiment. *Eval Program Plann* 1980;3(4):269–76.
- [34] De Allegri M, Louis V, Tiendrébeogo J, Souares A, Yé M, Tozan Y, et al. Moving towards universal coverage with malaria control interventions: achievements and challenges in rural Burkina Faso. *Int J Health Plann Manage* 2012. <http://dx.doi.org/10.1002/hpm.2116> [Epub ahead of print] PubMed PMID: 22689390.
- [35] Carroll C, Patterson M, Wood S, Booth A, Rick J, Balain S. A conceptual framework for implementation fidelity. *Implement Sci* 2007;2:40.

- [36] Perez D, Lefevre P, Castro M, Sanchez L, Toledo ME, Vanlerberghe V, et al. Process-oriented fidelity research assists in evaluation, adjustment and scaling-up of community-based interventions. *Health Policy Plan* 2011;26:413–22.
- [37] Patton MQ. *Utilization-focused evaluation*, 4th ed., Thousand Oaks, London, New Delhi: Sage Publications; 2008, 688 p.
- [38] Yin RK, Ridde V. Théorie et pratiques des études de cas en évaluation. In: Ridde V, Dagenais C, editors. *Approches et pratiques en évaluation de programme*. 2^e éd., Montréal: Presses de l'Université de Montréal; 2012 p. 177–93.
- [39] Nitiéma A, Ridde V, Girard JE. L'efficacité des politiques publiques de santé dans un pays de l'Afrique de l'Ouest : le cas du Burkina Faso. *Int Pol Sci Rev* 2003;24(2):237–56.
- [40] Ridde V, Robert E, Guichard A, Blaise P, Van Olmen J. L'approche Realist à l'épreuve du réel. *Rev Can Eval Programme* 2012;26(3): 37–59.
- [41] Astbury B, Leeuw FL. Unpacking black boxes: mechanisms and theory building in evaluation. *Am J Eval* 2010;31(3):363–81.
- [42] White H, Phillips D. Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework. New Delhi: International Initiative for Impact Evaluation; 2012.
- [43] Merton RK. *Social theory and social structure*. New York: Free Press; 1968 [xxiii, 702 p].
- [44] Pawson R, Tilley N. *Realistic evaluation*. London, Thousand Oaks (Cal): Sage Publications; 1997 [xvii, 235 p].
- [45] Kane S, Gerretsen B, Scherpbier R, Dal Poz M, Dieleman M. A realist synthesis of randomised control trials involving use of community health workers for delivering child health interventions in low and middle income countries. *BMC Health Serv Res* 2010;10(286).
- [46] Ridde V, Guichard A. Perception de quelques mécanismes favorables à la réduction des inégalités sociales de santé en France. *Global Health Promotion* 2011;18(3):47–60.