
18. Planifier la surveillance et l'évaluation des politiques

Atle Fretheim^{1}, Andrew D Oxman², John N Lavis³ et Simon Lewin⁴*

Fretheim A, Oxman AD, Lavis JN, Lewin S: **SUPPORT Tools for evidence-informed health Policymaking (STP). 18. Planning monitoring and evaluation of policies.** Health Research Policy and Systems; 2009, 7(Suppl 1):S18 doi:10.1186/1478-4505-7-S1-S18.
<http://www.health-policy-systems.com/content/pdf/1478-4505-7-S1-s18.pdf>

1. Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway; Section for International Health, Institute of General Practice and Community Medicine, Faculty of Medicine, University of Oslo, Norway
2. Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway
3. Centre for Health Economics and Policy Analysis, Department of Clinical Epidemiology and Biostatistics, and Department of Political Science, McMaster University, 1200 Main St. West, HSC-2D3, Hamilton, ON, Canada, L8N 3Z5
4. Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway; Health Systems Research Unit, Medical Research Council of South Africa

* Personne-ressource principale (atle.fretheim@nokc.no)

Il s'agit de la traduction d'un article publié dans le Health Research Policy and Systems, 2009; 7:Supplement 1 (www.health-policy-systems.com/supplements/7/S1). La reproduction, la distribution et l'utilisation de cet article, par quelque moyen que ce soit, sont permises à condition d'en citer la source. Le site Web SUPPORT (www.support-collaboration.org) renferme les hyperliens vers les versions chinoise, française, portugaise et espagnole. Vous pouvez envoyer vos commentaires sur la façon d'améliorer les outils présentés dans cette série de documents, par courriel, à : STP@nokc.no.

Cette série d'articles a été préparée dans le cadre du projet SUPPORT, parrainé par le programme INCO du sixième programme cadre de la Commission européenne, numéro de contrat 031939. L'Agence norvégienne de coopération pour le développement (NORAD), l'Alliance pour la recherche sur les politiques et les systèmes de santé et le Milbank Memorial Fund ont financé une réunion visant l'examen par des pairs d'une version initiale de la série. John Lavis a touché un salaire de la Chaire de recherche du Canada sur le transfert et l'échange des connaissances. NORAD, la composante norvégienne du groupe Cochrane Effective Practice and Organisation of Care (EPOC), le Centre norvégien de connaissances pour les services de santé, l'AHPSR, la Fondation canadienne de la recherche sur les services de santé (FCRSS), l'Evidence-Informed Policy Network (EVIPNet) et l'Organisation panaméricaine de la santé (OPS) ont participé à la traduction et à la diffusion des articles. Aucun des bailleurs de fonds n'a influé sur la rédaction, la révision ou l'approbation du contenu de la série.

La Fondation canadienne de la recherche sur les services de santé a traduit cet article en français avec l'appui du Centro Rosarino de Estudios Perinatales (CREP) et de l'Organisation panaméricaine de la santé (OPS).



Résumé

Le présent article fait partie d'une série de documents à l'intention des personnes chargées de prendre des décisions relativement aux politiques et aux programmes de santé ainsi que de celles qui les assistent.

Le terme *surveillance* sert généralement à décrire le processus qui consiste à recueillir des données de façon systématique pour que les responsables de politiques, les gestionnaires et d'autres parties intéressées puissent s'assurer qu'une nouvelle politique ou un nouveau programme est mis en œuvre conformément à leurs attentes. La surveillance repose sur des indicateurs qui permettent de juger, par exemple, si l'on a atteint les objectifs ou si les fonds alloués sont dépensés comme il se doit. Les termes *évaluation* et *surveillance* sont parfois utilisés de façon interchangeable, mais le premier est habituellement signe d'une plus grande préoccupation d'atteindre des résultats. Quant à l'expression *évaluation d'impact*, elle suppose, en règle générale, que l'on a pour objectif précis de déterminer si les changements observés sur le plan des résultats sont attribuables à une politique ou à un programme donné. Dans le présent article, nous proposons quatre questions pouvant orienter la surveillance et l'évaluation des propositions de politiques ou de programmes : 1. La surveillance s'impose-t-elle? 2. Qu'est-ce qui doit être mesuré? 3. Faut-il effectuer une évaluation d'impact? 4. Comment l'évaluation d'impact devrait-elle être effectuée?

À PROPOS DES OUTILS DU PROJET SUPPORT

Le présent article fait partie d'une série de documents destinés aux personnes chargées de prendre des décisions relativement aux politiques et aux programmes de santé et à celles qui les assistent, dans le but de les aider à s'assurer que leurs décisions sont bien éclairées par les meilleures données de recherche disponibles. L'introduction décrit plus en profondeur les outils du projet SUPPORT et les manières dont ils peuvent être utilisés [1]. Le glossaire de la série est joint à chaque article (voir le fichier complémentaire 1). Les résumés des études méthodiques préparés dans le cadre du projet SUPPORT peuvent être consultés à www.support-collaboration.org (en anglais). Des résumés en français, espagnol, portugais et chinois seront affichés sur ce site au cours de 2010 (www.support-collaboration.org/supporttool.htm). Toute rétroaction visant l'amélioration des outils abordés dans la série est la bienvenue et doit être envoyée à STP@nokc.no.

SCÉNARIOS

Scénario 1: Vous êtes un haut fonctionnaire responsable de la gestion globale de plusieurs programmes de services de santé. Vous désirez vous assurer que vous disposez de toute l'information nécessaire pour évaluer tant le rendement que l'impact de divers programmes.

Scénario 2: Vous travaillez au ministère de la Santé et vous avez été chargé de rédiger une note de service sur les diverses questions qu'il faudra prendre en considération lors de l'évaluation du programme national de vaccination.

Scénario 3: Vous travaillez au sein d'une unité autonome qui aide le gouvernement à mettre à profit les données probantes aux fins de l'élaboration de politiques et vous élaborez un plan de contrôle et d'évaluation du programme national de lutte contre la tuberculose.

CONTEXTE

Aux responsables de politiques (scénario 1), le présent article propose diverses questions à poser à leur personnel lors de la planification de la surveillance et de l'évaluation d'une nouvelle politique.

Aux personnes qui les assistent (scénarios 2 et 3), il suggère des questions à envisager au moment de planifier la surveillance de la mise en œuvre de politiques et de programmes ainsi que l'évaluation de leurs répercussions.

Les responsables de politiques et d'autres parties intéressées ont souvent besoin de savoir si une nouvelle politique ou un nouveau programme a été mis en œuvre conformément à leurs attentes. La mise en œuvre du programme se déroule-t-elle comme prévu? A-t-on atteint les objectifs et les fonds alloués sont-ils dépensés comme il se doit? Le terme *surveillance* sert généralement à décrire le processus qui consiste à recueillir des données de façon systématique en vue de trouver réponse à ces questions, alors que *surveillance du rendement* est l'expression souvent employée lorsqu'une évaluation a pour objectif principal de déterminer si la mise en œuvre d'un projet, d'un programme ou d'une politique permet d'obtenir les résultats escomptés [2].

Il est d'usage de faire appel à des *indicateurs* dans le cadre d'un processus de surveillance. Un indicateur se définit comme un facteur ou une variable de nature quantitative ou qualitative qui constitue un moyen simple et fiable de mesurer et de refléter des changements liés à une intervention ou d'aider à en apprécier le rendement [2]. Il peut s'agir d'un simple dénombrement d'événements (p. ex., le nombre de vaccinations effectuées au cours d'une période donnée) ou d'une hypothèse fondée sur diverses sources de données (p. ex., la proportion des enfants qui ont reçu tous les vaccins recommandés avant l'âge d'un an).

Les termes *évaluation* et *surveillance* sont parfois utilisés de façon interchangeable, mais le premier est habituellement signe d'une plus grande préoccupation d'atteindre des résultats. Cependant, leur emploi n'est pas uniforme et leur signification peut différer selon les personnes. Quant à l'expression *évaluation d'impact*, elle suppose, en règle générale, que l'on a pour objectif précis de déterminer si les changements observés sur le plan des résultats (ou impacts) sont attribuables à une politique ou à un programme donné.

QUESTIONS À ENVISAGER

1. La surveillance s'impose-t-elle?
2. Qu'est-ce qui doit être mesuré?
3. Faut-il effectuer une évaluation d'impact?
4. Comment l'évaluation d'impact devrait-elle être effectuée?

1. La surveillance s'impose-t-elle?

L'importance de la surveillance dépend de la mesure dans laquelle les parties intéressées estiment avoir besoin d'en savoir plus sur ce qui se passe « sur le terrain ».

Pour déterminer s'il convient d'établir un système pour surveiller une politique ou un programme, il faut tenir compte de plusieurs facteurs, notamment :

- S'il existe déjà un système de surveillance comprenant les indicateurs souhaités ou si de nouveaux indicateurs s'imposent

- Les coûts probables d'établissement du système requis (p. Ex., peut-on ajouter quelques nouveaux éléments aux procédures de collecte de données en vigueur ou bien faut-il effectuer d'autres enquêtes-ménages à grande échelle ou élaborer un tout nouvel outil?)
- Si les conclusions sont susceptibles d'être utiles (quelles mesures faudrait-il prendre si la surveillance révèle que les choses ne se passent pas comme prévu?)

La surveillance ne présente aucun intérêt si les données ne sont pas mises à profit. Celles-ci sont particulièrement utiles lorsqu'une mesure corrective est prise par suite de la découverte d'un écart entre les résultats escomptés et les résultats réels. Les attentes doivent souvent être réévaluées à la lumière de telles constatations, soit en vérifiant si les plans originaux n'étaient pas trop ambitieux, par exemple, ou si une nouvelle politique a fonctionné aussi efficacement que prévu.

Le tableau 1 présente deux exemples de systèmes de surveillance mis en place au sein de systèmes de santé [3,4].

2. Qu'est-ce qui doit être mesuré?

Pour surveiller la mise en œuvre d'une proposition de politique ou de programme, on fait appel, en règle générale, à des indicateurs qui portent sur différents éléments de la « chaîne de résultats » (c.-à-d. ressources, activités, réalisations, résultats ou impacts – voir la figure 1). Il est parfois suffisant de surveiller les ressources (le personnel ou l'équipement, par exemple), mais, dans d'autres cas, il peut être important de surveiller les activités du programme ou ses résultats (comme le nombre d'enfants complètement immunisés).

Un certain nombre de facteurs doivent présider au choix des indicateurs [5,6]:

- *La validité* : degré d'exactitude de la mesure qu'un indicateur est censé exprimer
- *L'acceptabilité* : mesure dans laquelle un indicateur est jugé acceptable par les personnes évaluées et celles qui procèdent à l'évaluation
- *La faisabilité* : mesure dans laquelle des données valides, fiables et convergentes peuvent être recueillies
- *La fiabilité* : ordre de grandeur des erreurs négligeables attribuables aux indicateurs, le cas échéant, ou possibilité de reproduire les conclusions si un autre organisme devait recueillir les mêmes données
- *La réactivité au changement* : mesure dans laquelle l'indicateur peut déceler les changements dans l'unité de mesure
- *La validité prédictive* : degré d'exactitude des résultats pertinents prédits par un indicateur

Les coûts liés à la collecte de données ainsi que la capacité d'analyser les données et de les transmettre aux gestionnaires et aux prestataires de services de santé peuvent également limiter le choix des indicateurs. Dans les milieux où les ressources

analytiques sont rares, il est sans doute préférable de choisir un indicateur simple, même s'il n'offre pas la meilleure validité prédictive, plutôt qu'un indicateur qui nécessite des manipulations statistiques.

Un compromis se dessine souvent entre, d'une part, la volonté d'utiliser les indicateurs optimaux souhaités et, d'autre part, l'obligation d'utiliser des indicateurs fondés sur les données existantes. Il y a de bonnes raisons de ne pas choisir plus d'indicateurs qu'il n'en faut essentiellement, notamment la nécessité de limiter le fardeau de la collecte des données au sein d'un système de santé, d'éviter la collecte de données qui ne seront pas utilisées et de se consacrer à la collecte de données de plus grande qualité, même si de la sorte, l'ensemble des données recueillies est plus restreint [7].

Les données courantes recueillies dans les systèmes de santé peuvent fournir de précieux renseignements et servir à la surveillance. La collecte des données peut aussi être expressément axée sur la surveillance (par enquêtes ou entrevues, par exemple). Il faut également tenir compte du degré de motivation des personnes censées recueillir les données. Souvent, les travailleurs de la santé devront intégrer cette tâche à un horaire journalier déjà chargé. En conséquence, s'ils estiment que l'information recueillie revêt peu d'intérêt pour eux ou pour la collectivité locale, ils ne seront guère motivés à s'en acquitter. De même, si des incitatifs ou des sanctions sont associés aux conclusions du processus de surveillance (p. ex., lorsque la rémunération des prestataires de services de santé est lié à des indicateurs de rendement), le risque de simulation et de manipulation des données ou du système ne peut être écarté.

3. Faut-il effectuer une évaluation d'impact?

Comme nous l'avons vu, l'une des limites des activités de surveillance est le fait qu'elles ne permettent pas nécessairement de savoir si une politique ou un programme a eu un impact sur les indicateurs mesurés. En effet, presque toujours, des facteurs autres que ceux associés à des interventions particulières influent sur les indicateurs employés à des fins de surveillance, de sorte qu'il est extrêmement difficile de déterminer quels facteurs sont à l'origine des changements observés. Si la surveillance révèle que le rendement s'améliore, cela ne signifie pas forcément que l'intervention en est le (seul) facteur de causalité. Il est possible que les indicateurs se soient améliorés de toute façon, même sans l'intervention (voir la figure 2).

L'établissement d'une relation de cause à effet entre un programme ou une politique et des changements dans les résultats est la pierre angulaire de l'évaluation d'impact. Selon la Banque mondiale, ce qui serait arrivé aux personnes qui ont fait l'objet d'une intervention en l'absence de celle-ci constitue véritablement *la* question centrale d'une évaluation de l'impact [8].

On a parfois d'excellentes raisons d'escompter des résultats positifs sur la foi d'une documentation solide issue, par exemple, d'évaluations antérieures. Cependant, il est

rare qu'on dispose de telles données probantes, ou alors les données disponibles ne peuvent s'appliquer au milieu d'intérêt. Il y a donc un risque réel qu'un nouveau programme soit inefficace ou, pire encore, qu'il cause plus de mal que de bien. C'est là une question que les responsables de politiques doivent absolument éclaircir lorsqu'ils mettent en œuvre de nouveaux programmes, d'autant plus que cette information pourrait être utile pour l'élaboration d'autres politiques, dans le même milieu ou ailleurs.

Les évaluations d'impact peuvent être coûteuses. On peut s'assurer que l'investissement est rentable en comparant les conséquences des deux options : procéder à l'évaluation ou s'en abstenir. Par exemple, selon toute probabilité, le programme sera-t-il interrompu ou modifié si les résultats se révélaient négatifs? Si la réponse est « non », l'intérêt d'effectuer une évaluation d'impact est manifestement limité.

En règle générale, une évaluation d'impact sera vraisemblablement plus rentable si on peut dégager des résultats à mesure que l'intervention se déroule. Il est alors possible de l'améliorer ou d'y mettre fin, selon les résultats de l'évaluation d'impact effectuée au tout début de la mise en œuvre. On peut donc optimiser les ressources selon deux éventualités : soit lorsqu'il est impossible de réaliser une étude pilote, soit lorsqu'il est possible et pratique de modifier ou d'interrompre l'intervention (s'il y a lieu) en fonction des résultats.

Le régime d'assurance maladie du gouvernement mexicain, *Seguro Popular*, constitue un bon exemple de l'intégration d'une évaluation d'impact à la mise en œuvre d'un programme [9-11]. Implanté en 2001, ce régime a été créé pour offrir la couverture d'assurance maladie à près de 50 millions de Mexicains qui ne pouvaient pas encore se prévaloir des programmes en vigueur. Ainsi, le gouvernement a profité du calendrier de la mise en œuvre progressive pour procéder à une comparaison des résultats dans les collectivités bénéficiant du régime à ceux observés dans les groupes toujours en attente. En plus de déterminer si la réforme avait permis d'atteindre les résultats escomptés sans avoir d'effets négatifs imprévus, l'évaluation a aussi été une occasion d'apprentissage partagé.

Une évaluation d'impact peut également être utile à l'issue de la mise en œuvre d'un programme, par exemple lorsque l'avenir de celui-ci est incertain. Par exemple, en vertu du programme mexicain de transferts conditionnels en espèces *Progresa* (rebaptisé *Oportunidades*), instauré au milieu des années 1990, des fonds sont versés aux familles bénéficiaires sous réserve du respect du principe de coresponsabilité (envoyer les enfants à l'école au lieu de les faire travailler, administrer des suppléments nutritionnels spécialement conçus pour les enfants en bas âge, fréquenter une clinique médicale où sont prévues à leur intention des interventions de promotion de la santé et de prévention des maladies [12]). Aux fins de l'évaluation, 506 collectivités ont été désignées au hasard pour participer au programme, soit immédiatement ou deux ans plus tard [13]. Les conclusions de cette évaluation d'impact ont permis d'éclairer

directement les décisions stratégiques du gouvernement mexicain et l'ont convaincu non seulement de maintenir le programme, mais aussi de l'élargir [12].

4. Comment l'évaluation d'impact devrait-elle être effectuée?

L'attribution d'un changement observé à une politique ou à un programme exige une comparaison entre les personnes ou les groupes touchés et d'autres qui ne le sont pas. Il importe aussi que les groupes comparés soient aussi semblables que possible afin d'éliminer toute influence autre que le programme en soi. C'est ce à quoi servent les *essais cliniques à répartition aléatoire* dans le cadre desquels les personnes ou les groupes qui participent au programme et ceux qui n'y participent pas sont répartis au hasard (par secteur géographique, par exemple). De tels essais prennent habituellement la forme de projets pilotes précédant la mise en œuvre d'un programme à l'échelle nationale. Ils peuvent aussi être menés parallèlement à la mise en œuvre en vraie grandeur, comme l'illustrent les exemples mexicains présentés ci-dessus.

Cependant, il n'est pas toujours possible de procéder à des essais de ce genre. Il faut alors comparer les changements avant et après la mise en œuvre du programme, auquel cas les changements observés doivent être survenus durant la même période dans les régions où le programme n'a pas été mis en œuvre (p. ex., dans les districts ou les pays voisins). C'est ce qu'on appelle une *étude comparative avant-après*. Ou encore, on recourra à une *analyse de série chronologique interrompue* pour laquelle les données sont recueillies à différents moments, avant, durant et après la mise en œuvre du programme.

En général, il n'est pas recommandé de se contenter de comparer la valeur d'un indicateur avant et après la mise en œuvre du programme, car le risque est grand de tirer des conclusions trompeuses. Ainsi, les changements observés (l'incidence du VIH, par exemple) peuvent résulter de facteurs connus ou inconnus autres que ceux liés au programme comme tel (voir la figure 2 [14,15]).

Le fichier complémentaire 2, à la fin du présent article, donne un aperçu de diverses méthodes d'évaluation. En outre, les forces et les faiblesses de chacune de ces méthodes sont présentées brièvement dans le fichier complémentaire 3.

Les évaluations d'impact devraient être planifiées bien avant la mise en œuvre du programme, de concert avec les parties intéressées, y compris les responsables de politiques. Une fois qu'un programme a été mis en œuvre à grande échelle, il est généralement trop tard pour évaluer les caractéristiques de référence ou former des groupes témoins adéquats. Ainsi, l'assignation aléatoire pour déterminer quelles collectivités participeront au programme ne peut se faire après la mise en œuvre du programme à l'échelle nationale. Les évaluations d'impact intégrées au programme dès le départ sont donc plus susceptibles de produire des conclusions valables que les évaluations menées après coup. En outre, si les évaluations d'impact sont perçues

comme faisant partie intégrante de la mise en œuvre des programmes, il est probable que les responsables de politiques et d'autres parties intéressées auront plus à cœur de tenir compte de leurs conclusions.

Le nombre de personnes ou de collectivités requises aux fins de l'évaluation d'impact devrait aussi être estimé au début du processus. On s'assurera ainsi de disposer d'un échantillon assez vaste pour tirer des conclusions significatives des constatations de l'évaluation.

Dans la sphère de la santé, comme dans tout autre domaine, les programmes doivent être à la fois efficaces et rentables. Pour évaluer les aspects économiques d'un programme, il faut estimer l'utilisation des ressources et les coûts, de préférence selon les données recueillies au cours d'une mise en œuvre réelle [16]. Les décisions quant au type de données économiques à recueillir devraient donc être prises aussi au début du processus, avant d'entreprendre l'évaluation.

Les évaluations d'impact sont susceptibles d'être plus profitables lorsqu'elles comportent une évaluation des processus permettant de vérifier si la proposition de politique ou de programme a été mise en œuvre tel que prévu. Ce type d'évaluation portera aussi sur les processus de mise en œuvre et de changement, les réactions au programme et approfondir les fondements des conclusions de l'évaluation [17].

Voir les exemples d'évaluations d'impact au tableau 2.

Les contraintes budgétaires, les échéanciers et les limites des données peuvent avoir un effet dissuasif et freiner les efforts pour assurer une mise en œuvre rigoureuse. Ces impératifs peuvent compromettre la fiabilité des évaluations d'impact, notamment :

- En réduisant la validité globale des résultats, par exemple, en raison d'une planification ou d'un suivi insuffisants, d'un manque de données de référence, du recours à des sources de données inadéquates ou du choix de groupes témoins inappropriés
- Par l'utilisation d'échantillons inadéquats, par exemple, parce qu'ils sont faciles à déterminer mais non représentatifs, parce que leur taille est insuffisante ou parce qu'on n'a pas assez tenu compte des facteurs contextuels

Il est possible de surmonter de telles difficultés en planifiant à l'avance ou en trouvant un moyen de réduire le coût de la collecte de données. Cependant, il importe de s'assurer que ni les facteurs susceptibles de nuire à la validité des résultats, ni les limites de l'échantillon ne sont tels que les résultats de l'évaluation ne puissent fournir d'information fiable. Avant de procéder à une évaluation, il faut donc d'abord s'assurer qu'une évaluation adéquate est possible. Sinon, il faut vérifier si le programme peut être mis en œuvre sans évaluation préalable, compte tenu de l'incertitude entourant ses répercussions éventuelles [18].

Les évaluations d'impact ne sont d'aucune utilité si les conclusions ne sont pas mises à profit. Les résultats devraient servir à éclairer les décisions sur la pertinence de poursuivre, de modifier ou d'interrompre les programmes en vigueur. Manifestement, d'autres intérêts devront aussi être pris en compte. Par exemple, les responsables de politiques pourraient décider de mettre en veilleuse certaines conclusions d'évaluations qui seraient incompatibles avec d'autres intérêts qu'ils jugent plus importants [19]. Il importe toutefois de ne pas exclure certaines conclusions d'évaluations d'impact, notamment pour des raisons politiques, car cela va à l'encontre d'un des principaux objectifs des évaluations d'impact, à savoir tirer des enseignements de l'expérience et partager les connaissances acquises. Le recours à une tierce partie indépendante pour réaliser les évaluations d'impact peut minimiser les risques de manipulation ou de dissimulation des constatations.

CONCLUSION

Le présent article décrit divers aspects de la surveillance et de l'évaluation. À l'heure actuelle, les activités de surveillance et d'évaluation sont souvent effectuées à l'aide de méthodes qui ne donnent pas d'évaluations valables de la mise en œuvre d'une politique ou d'un programme, ni d'estimations valables des effets. Il arrive que l'on passe outre toute forme d'évaluation. En tenant compte des éléments exposés précédemment, les responsables de politiques et les personnes qui les assistent devraient être en mesure d'élaborer des plans qui généreront des connaissances nouvelles et directement utilisables.

RESSOURCES

Documents utiles et lecture complémentaire

Segone M (dir.). 2008. Bridging the gap: The role of monitoring and evaluation in evidence-based policy making. UNICEF, Banque mondiale et International Development Evaluation Association.

www.unicef.org/ceecis/evidence_based_policy_making.pdf

MacKay K. 2007. How to Build M&E Systems to Support Better Government. Groupe d'évaluation indépendante (Banque mondiale).

http://www.worldbank.org/oed/ecd/better_government_fr.html

Groupe d'évaluation indépendante (Banque mondiale). 2004. Monitoring and Evaluation (M&E): Some Tools, Methods and Approaches.

[http://lnweb90.worldbank.org/oed/oeddoelib.nsf/24cc3bb1f94ae11c85256808006a0046/a5efbb5d776b67d285256b1e0079c9a3/\\$FILE/MandE_tools_methods_approaches.pdf](http://lnweb90.worldbank.org/oed/oeddoelib.nsf/24cc3bb1f94ae11c85256808006a0046/a5efbb5d776b67d285256b1e0079c9a3/$FILE/MandE_tools_methods_approaches.pdf)

National Treasury of South Africa. 2007. Framework for Managing Programme Performance Information. www.treasury.gov.za/publications/guidelines/FMPI.pdf

Barber S. 2007. Health system strengthening interventions: Making the case for impact evaluation – Briefing note 2. Genève, Alliance pour la recherche sur les politiques et les systèmes de santé.

www.who.int/alliance-hpsr/resources/Alliance%20%20HPSR%20-%20Briefing%20Note%202.pdf

Savedoff WD, Levine R, Birdsall N. 2006. « When will we ever learn? Improving lives through impact evaluation ». Report of the Evaluation Gap Working Group. Center for Global Development. www.cgdev.org/content/publications/detail/7973/

Grimshaw J, Campbell M, Eccles M, Steen N. 2000. « Experimental and quasi-experimental designs for evaluating guideline implementation strategies ». Family Practice; 17: S11-S18. http://fampra.oxfordjournals.org/cgi/reprint/17/suppl_1/S11

Sites Web

Groupe d'évaluation indépendante (Independent Evaluation Group – IEG) de la Banque mondiale : www.banquemondiale.org/ieg/about.htm – Le Groupe d'évaluation indépendante est un organe interne de la Banque mondiale mais autonome. Il détermine si les actions entreprises sont couronnées ou non de succès, évalue les mesures adoptées par les emprunteurs pour exécuter un projet et en assurer la pérennité, et mesure la contribution de la Banque au développement général d'un pays à long terme

International Initiative for Impact Evaluation (3ie) : www.3ieimpact.org – L'organisme, qui a pour but d'améliorer la vie des groupes défavorisés dans les PRFI, fournit des données probantes (et leur synthèse) sur les propositions de politiques efficaces, sur les facteurs d'efficacité et le moment où elle s'exerce, ainsi que sur les coûts afférents

Health Metrics Network : www.who.int/healthmetrics/en – Le réseau Health Metrics Network (HMN) a pour objectif d'accroître la disponibilité et l'utilisation en temps opportun d'une information en matière de santé rigoureuse. À cette fin, HMN établit des stratégies pour la mise en place et le renforcement des systèmes d'information sur la santé (SIS), il soutient les pays dans la mise en œuvre de réformes des SIS et diffuse des connaissances sur les biens publics mondiaux par la recherche, l'innovation technique et la mise en commun des enseignements retenus

NorthStar : www.rebeqi.org/?pageID=34&ItemID=35 – NorthStar est un outil servant à planifier des programmes d'amélioration de la qualité, à les mettre en œuvre et à les évaluer

CONFLIT D'INTÉRÊTS

Les auteurs déclarent n'avoir aucun conflit d'intérêts.

CONTRIBUTION DE CHAQUE AUTEUR

AF a rédigé l'ébauche initiale du présent article; ADO, JNL et SL ont contribué à ce travail en plus de réviser le document.

REMERCIEMENTS

L'introduction de la présente série comprend nos remerciements aux bailleurs de fonds et aux personnes qui y ont contribué. Nous tenons toutefois à exprimer notre reconnaissance particulière à Ruth Levine pour ses commentaires avisés sur une version antérieure du présent article.

RÉFÉRENCES BIBLIOGRAPHIQUES

1. Lavis JN, Oxman AD, Lewin S, Fretheim A: **SUPPORT Tools for evidence-informed health Policymaking (STP). Introduction.** *Health Res Policy Syst* 2009, **7 (Suppl 1:I1)**.
2. Development Assistance Committee Working Party on Aid Evaluation: *Glossary of Key Terms in Evaluation and Results Based Management*. Paris, OECD Publications. 2002.
3. Harries AD, Gomani P, Teck R, de Teck OA, Bakali E, Zachariah R, et al: **Monitoring the response to antiretroviral therapy in resource-poor settings: the Malawi model.** *Trans R Soc Trop Med Hyg* 2004, **98**:695-701.
4. Jakobsen E, Palshof T, Osterlind K, Pilegaard H: **Data from a national lung cancer registry contributes to improve outcome and quality of surgery: Danish results.** *Eur J Cardiothorac Surg* 2009, **35**:348-52.
5. Smith PC, Mossialos E, Papanicolas I: *Performance measurement for health system improvement: experiences, challenges and prospects*. Background Document for WHO European Ministerial Conference on Health Systems: "Health Systems, Health and Wealth". Copenhagen, World Health Organization, Europe. 2008.
6. Campbell SM, Braspenning J, Hutchinson A, Marshall M: **Research methods used in developing and applying quality indicators in primary care.** *Qual Saf Health Care* 2002, **11**:358-64.
7. MacKay K: *How to Build M&E Systems to Support Better Government*. Washington DC, The World Bank. 2007.

8. The World Bank: *Impact evaluation: Overview*. [<http://go.worldbank.org/2DHMCRFFT2>]. The World Bank. 2009.
9. Moynihan R, Oxman A, Lavis JN, Paulsen E: *Evidence-Informed Health Policy: Using Research to Make Health Systems Healthier*. Rapport nr. 1-2008. Oslo, Nasjonalt kunnskapssenter for helsetjenesten. 2008.
10. Frenk J, Gonzalez-Pier E, Gomez-Dantes O, Lezana MA, Knaul FM: **Comprehensive reform to improve health system performance in Mexico**. *Lancet* 2006, **368**:1524-34.
11. Gakidou E, Lozano R, Gonzalez-Pier E, Abbott-Klafter J, Barofsky JT, Bryson-Cahn C, et al: **Assessing the effect of the 2001-06 Mexican health reform: an interim report card**. *Lancet* 2006, **368**:1920-35.
12. Frenk J: **Bridging the divide: global lessons from evidence-based health policy in Mexico**. *Lancet* 2006, **368**:954-61.
13. Rivera JA, Sotres-Alvarez D, Habicht JP, Shamah T, Villalpando S: **Impact of the Mexican program for education, health, and nutrition (Progresá) on rates of growth and anemia in infants and young children: a randomized effectiveness study**. *JAMA* 2004, **291**:2563-70.
14. Savedoff WD, Levine R, Birdsall N: *When will we ever learn? Improving lives through impact evaluation*. Washington DC, Center for Global Development. 2006.
15. Shadish WR, Cook TD, Campbell DT: *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin; 2002.
16. Oxman AD, Fretheim A, Lavis JN, Lewin S: **SUPPORT Tools for evidence-informed health Policymaking (STP). 12. Finding and using research evidence about resource use and costs**. *Health Res Policy Syst* 2009, **7** (Suppl 1:S12).
17. Lewin S, Glenton C, Oxman AD: **Use of qualitative methods alongside randomised controlled trials of complex healthcare interventions: methodological study**. *BMJ* 2009, **339**:b3496.
18. Oxman AD, Lavis JN, Fretheim A, Lewin S: **SUPPORT Tools for evidence-informed health Policymaking (STP). 17. Dealing with insufficient research evidence**. *Health Res Policy Syst* 2009, **7** (Suppl 1:S17).
19. Scheel IB, Hagen KB, Oxman AD: **The unbearable lightness of healthcare policy making: a description of a process aimed at giving it some weight**. *J Epidemiol Community Health* 2003, **57**:483-7.
20. Amuron B, Coutinho A, Grosskurth H, Nabiryo C, Birungi J, Namara G, et al: **A cluster-randomised trial to compare home-based with health facility-based antiretroviral treatment in Uganda: study design and baseline findings**. *Open AIDS J* 2007, **1**:21-7.
21. Jaffar S, Amuron B, Birungi J, Namara G, Nabiryo C, Coutinho A, et al: **Integrating research into routine service delivery in an antiretroviral treatment programme: lessons learnt from a cluster randomized trial comparing strategies of HIV care in Jinja, Uganda**. *Trop Med Int Health* 2008, **13**:795-800.
22. London School of Hygiene and Tropical Evidence: *Home-based HIV care just as effective as clinic-based care in Sub-saharan Africa*.

[\[www.lshtm.ac.uk/news/2009/homeHIVcare.html\]](http://www.lshtm.ac.uk/news/2009/homeHIVcare.html). London School of Hygiene and Tropical Evidence, University of London. 2009.

23. Fretheim A, Havelrud K, MacLennan G, Kristoffersen DT, Oxman AD: **The effects of mandatory prescribing of thiazides for newly treated, uncomplicated hypertension: interrupted time-series analysis.** *PLoS Med* 2007, **4**:e232.

Figure 1. Modèle de chaîne de résultats (définitions adaptées du Glossaire de l'OCDE [2])

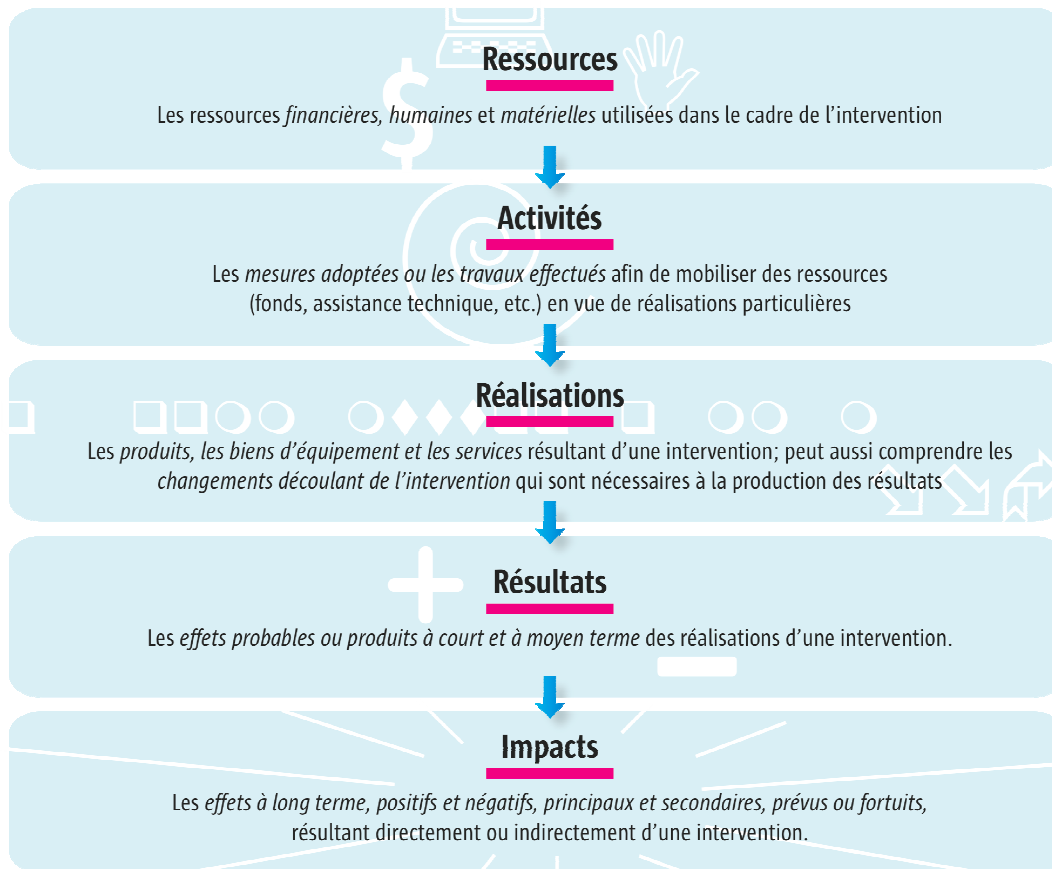
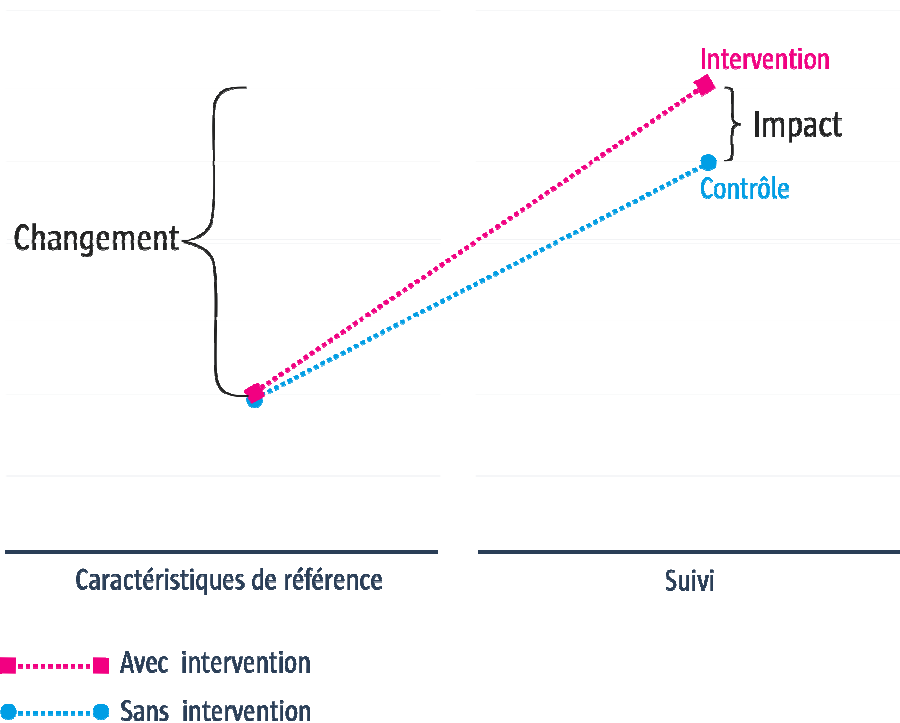


Figure 2. Changements liés au rendement avec et sans intervention*



* Il peut être trompeur de considérer que le changement survenu par suite d'une intervention est attribuable non plus aux « caractéristiques de référence » mais au « suivi », notamment parce que, en l'occurrence, il y a eu amélioration aussi sur le plan du « contrôle ». Même en ce cas, rien n'indique que l'on puisse attribuer avec certitude la différence entre l'intervention et le contrôle au programme ou à l'intervention, d'autres variations à cet égard pouvant être à l'origine des écarts observés relativement à l'indicateur étalon

Tableau 1. Exemples de méthodes de surveillance au sein de systèmes de santé

Généralisation de l'accès à la thérapie antirétrovirale (TAR) au Malawi [3]

Lorsque les autorités sanitaires du Malawi ont décidé de mettre la thérapie antirétrovirale à la portée d'une grande proportion de la population séropositive, elles ont mis sur pied un système de surveillance de la mise en œuvre de cette nouvelle politique. Ce système s'inspire de la méthode utilisée par l'OMS pour surveiller les programmes nationaux de lutte contre la tuberculose. Chaque patient qui entreprend la thérapie reçoit une carte d'identité portant un numéro unique qui est conservée à la clinique. L'information recueillie auprès des nouveaux patients comprend leurs nom, adresse, âge et taille, le nom de leur tuteur ainsi que la raison pour laquelle ils entreprennent la thérapie. On demande aux patients de se rendre à la clinique une fois par mois pour obtenir leurs médicaments. Au cours de cette visite, on enregistre leur poids et on leur pose des questions sur leur état général, leur statut ambulatoire, leur travail et tout effet secondaire des médicaments. On veille aussi au dénombrement des doses restantes afin de s'assurer de l'observance thérapeutique. De plus, les autorités procèdent à la consignation mensuelle de résultats normalisés selon les catégories suivantes :

- *Vivant* : le patient est vivant et il est venu chercher son approvisionnement de 30 jours de médicaments
- *Décédé* : le patient est décédé durant le traitement
- *Perdu de vue* : le patient ne s'est pas présenté depuis trois mois
- *Arrêt* : le patient a cessé le traitement à cause des effets secondaires ou pour d'autres raisons
- *Transfert* : le patient reçoit désormais un autre traitement

Les données recueillies dans le cadre du système malawien de surveillance de la distribution des traitements antirétroviraux peuvent être analysées et utilisées de diverses façons. La comparaison peut porter sur les résultats des traitements administrés à des patients recrutés à divers moments. Par exemple, l'augmentation du taux d'abandon d'un traitement de premier recours au profit d'une thérapie de deuxième intention, ou du taux de mortalité, pourrait être attribuable à la pharmacorésistance dans le traitement de première intention. À l'inverse, la diminution du nombre de décès ou de patients perdus de vue indiquerait que la gestion du programme de thérapie antirétrovirale s'améliore. Si les résultats sont particulièrement faibles dans certaines régions géographiques ou certaines cliniques, des mesures pourraient s'imposer pour faire face à cette situation.

Chirurgie du cancer du poumon au Danemark [4]

Les autorités danoises ont émis des directives de pratique clinique en matière de gestion du cancer du poumon par suite des résultats défavorables chez les patients traités par la chirurgie. Pour surveiller la mise en œuvre des directives, elles ont établi un registre national des cas de cancer du poumon, lequel comporte des renseignements précis au sujet des patients traités par la chirurgie. Les indicateurs sélectionnés aux fins de ce registre (Danish Lung Cancer Registry) portent notamment sur l'étendue du cancer dans l'organisme (ou stade), l'intervention chirurgicale utilisée, toute complication survenue et les taux de survie.

Les données du registre danois servent notamment à s'assurer du respect des lignes directrices nationales en matière de chirurgie du cancer du poumon. Des vérifications sont effectuées à l'échelle locale, régionale et nationale afin de cerner les problèmes ou les facteurs susceptibles de faire obstacle à l'observance des lignes directrices et leurs conclusions servent de fondement à l'élaboration de stratégies d'amélioration de la qualité.

Tableau 2. Exemples d'évaluations d'impact

Traitement antirétroviral à domicile en Ouganda [20-22]

La pénurie de personnel clinique et les difficultés d'accès aux soins en raison des coûts de transport constituent un obstacle majeur à la distribution générale de traitements antirétroviraux dans les pays en développement. L'une des solutions proposées est le traitement à domicile où la distribution des médicaments, la surveillance de l'état de santé et le soutien aux patients sont assurés au domicile du patient par un personnel non qualifié sur le plan clinique. Il est loin d'être certain toutefois que cette stratégie permette d'offrir des soins d'une qualité acceptable – y compris des orientations médicales en temps opportun – ou qu'un tel système soit rentable. En conséquence, avant d'implanter des programmes de soins à domicile sur une grande échelle, il importe d'en évaluer le rapport coût-efficacité.

Pour assurer une juste comparaison entre le traitement antirétroviral à domicile et en clinique, les chercheurs d'Ouganda ont mené un essai clinique à répartition aléatoire. La zone d'étude a été divisée en 44 sous-zones géographiques distinctes. Des soins à domicile ont été offerts dans certains secteurs, tandis que dans les autres les services de santé en établissement se sont poursuivis. La répartition des secteurs devant bénéficier de l'un ou l'autre des systèmes a été effectuée de façon aléatoire, de manière à minimiser la possibilité que de grandes différences entre les groupes de référence puissent influencer sur les résultats si, par exemple, la décision d'offrir les soins à domicile revenait aux districts ou si elle était fondée sur leur capacité de dispenser ces services à domicile. La répartition aléatoire était aussi la façon la plus équitable de décider où implanter le nouveau système, puisque chaque district avait ainsi une chance égale d'être choisi.

Les chercheurs ont fait valoir que les soins à domicile prodigués par des travailleurs de la santé non professionnels étaient aussi efficaces que les services de santé offerts en clinique et dirigés par une infirmière ou un médecin.

Utilisation obligatoire des diurétiques thiazidiques dans la prise en charge de l'hypertension artérielle en Norvège [23]

Afin de limiter les coûts, les responsables de politiques de la Norvège ont décidé que l'on prescrirait dorénavant des diurétiques thiazidiques pour le traitement de l'hypertension artérielle au lieu d'autres médicaments plus coûteux, dans les cas où ceux-ci devaient être remboursés. La politique a été mise en œuvre à l'échelle nationale quelques mois après cette décision. Les détracteurs faisant valoir que la nouvelle politique ne donnerait pas les résultats espérés, le ministère de la Santé a parrainé une étude en vue d'évaluer l'impact de la politique qu'il avait mise en œuvre.

La prescription obligatoire de diurétiques thiazidiques pour le traitement de l'hypertension artérielle a été implantée dans l'ensemble de la Norvège avec une telle rapidité qu'il était impossible d'effectuer une évaluation d'impact rigoureuse et planifiée. Cependant, les dossiers médicaux électroniques de 61 cliniques ont permis aux chercheurs d'extraire, ultérieurement des données sur les médicaments prescrits un an avant et un an après l'instauration de la nouvelle politique. Ils ont eu recours à l'analyse de série chronologique interrompue pour déterminer, entre autres résultats d'intérêt, l'évolution au fil du temps des taux mensuels de prescription de diurétiques thiazidiques et voir si des changements significatifs pourraient être attribuables à la politique mise en œuvre. L'analyse a fait ressortir une augmentation importante du recours aux diurétiques thiazidiques (de 10 à 25 % durant la période de transition de trois mois préétablie), à la suite de quoi l'utilisation du médicament s'est stabilisée.

Fichier complémentaire 2. Méthodes d'évaluation (adaptation de Cochrane Handbook for Systematic Reviews of Interventions*)

Essai clinique à répartition aléatoire	<ul style="list-style-type: none">• Étude expérimentale dans laquelle des personnes sont sélectionnées au hasard pour faire l'objet de différentes interventions (p. ex., en tirant à pile ou face ou en générant une liste de nombres aléatoires par ordinateur)
Essai clinique à répartition aléatoire en grappes	<ul style="list-style-type: none">• Étude expérimentale dans laquelle des groupes de personnes (p. ex., des classes ou des hôpitaux) sont sélectionnés au hasard pour faire l'objet de différentes interventions
Essai clinique non aléatoire	<ul style="list-style-type: none">• Étude expérimentale dans laquelle des personnes sont sélectionnées pour faire l'objet de différentes interventions par des méthodes non aléatoires (p. ex., les patients admis durant la Semaine 1 reçoivent l'intervention A, ceux de la Semaine 2, l'intervention B, et ainsi de suite)
Étude comparative avant- après	<ul style="list-style-type: none">• Étude au cours de laquelle des observations sont faites avant et après la mise en œuvre d'une intervention, tant dans le groupe qui fait l'objet de l'intervention que dans le groupe témoin qui, lui, ne fait l'objet d'aucune intervention. La collecte de données devrait habituellement se faire simultanément dans les deux groupes
Étude de série chronologique interrompue	<ul style="list-style-type: none">• Étude où l'on fait des observations à différents moments avant et après une intervention. Les mesures sont <i>interrompues</i> par l'intervention. Cette méthode vise à déterminer si une intervention donnée a eu un effet significativement plus important que toute tendance sous-jacente au fil du temps
Étude comparative avec groupe témoin historique	<ul style="list-style-type: none">• Étude comparant un groupe de participants qui font l'objet d'une intervention avec un groupe semblable formé antérieurement et qui n'avait pas fait l'objet d'une telle intervention
Étude de cohorte	<ul style="list-style-type: none">• Étude dans laquelle un groupe déterminé de personnes (la cohorte) est suivi au fil du temps afin d'étudier les corrélations entre les diverses interventions dont les participants ont fait l'objet et les résultats subséquents. Dans une étude de cohortes <i>prospective</i>, les participants

sont recrutés avant toute intervention et sont suivis par la suite. Dans une étude de cohortes *rétrospective*, les données prévalant au choix des sujets sont tirées de dossiers antérieurs où sont décrites les interventions dont ils ont fait l'objet, et ces sujets sont suivis rétrospectivement à partir du moment où les dossiers ont été constitués

- Étude cas-témoin**
- Étude comparant des personnes qui ont un résultat d'intérêt particulier (cas) à des gens de la même population source mais sans ce résultat (témoin), afin d'examiner la corrélation entre le résultat et l'exposition préalable (p. ex., l'intervention dont ils ont fait l'objet). Cette méthode est particulièrement utile lorsque le résultat est rare
-

- Étude transversale**
- Étude dans laquelle on recueille de l'information sur des interventions passées ou courantes et sur des résultats de santé courants pour un groupe de personnes, à un moment précis. Ce genre d'étude porte sur les corrélations entre les résultats et l'exposition à des interventions
-

- Étude qualitative**
- Étude menée dans un cadre naturel généralement en vue d'interpréter ou de des phénomènes selon la signification que les gens leur donnent. Habituellement, dans une étude de ce genre, les chercheurs recueillent, auprès de personnes ou de groupes d'« informants », ou en consultant des documents, des données narratives qu'ils interprètent par la suite
-

* Cochrane Collaboration. 2002. Cochrane Handbook for Systematic Reviews of Interventions. Chichester: The Cochrane Collaboration et John Wiley & Sons Ltd.

Fichier complémentaire 3. Forces et faiblesses des méthodes d'évaluation : quelques exemples

	Forces	Faiblesses
Essai clinique à répartition aléatoire	<ul style="list-style-type: none"> • Méthode généralement considérée comme la plus solide pour établir des relations de cause à effet, ce qui est la pierre angulaire de l'évaluation d'impact 	<ul style="list-style-type: none"> • Peut exiger beaucoup de temps et présenter des défis sur le plan logistique • Les résultats ne sont pas nécessairement transférables à d'autres milieux
Essai clinique à répartition aléatoire en grappes	<ul style="list-style-type: none"> • Même points forts que les essais cliniques à répartition aléatoire ordinaires. En outre, le risque de « contamination » est moindre (p. ex., le risque que l'intervention A soit administrée aux personnes censées faire l'objet de l'intervention B seulement ou qu'elle leur soit préjudiciable). Ainsi, lorsque des infirmières sont sélectionnées au hasard pour mettre en œuvre un nouveau programme, d'autres infirmières peuvent être influencées par ces changements et décider de se livrer aux mêmes activités. Il serait alors préférable de faire une sélection aléatoire des services, et d'inclure l'ensemble de leur personnel, plutôt que des infirmières en particulier 	<ul style="list-style-type: none"> • Les différences quant aux situations de référence peuvent poser problème étant donné que le nombre d'unités (ou de <i>grappes</i>) sélectionnées au hasard est généralement moindre que dans un essai où ce sont des personnes qui sont réparties de façon aléatoire. Peut exiger beaucoup de temps et présenter des défis sur le plan logistique, mais moins qu'un essai clinique à répartition aléatoire ordinaire

	Forces	Faiblesses
Essai clinique non aléatoire	<ul style="list-style-type: none"> • Peut être plus facile et plus pratique à réaliser qu'un essai à répartition aléatoire 	<ul style="list-style-type: none"> • Lorsque l'assignation n'est effectuée par des méthodes aléatoires, des biais de sélection peuvent se produire, p. ex., parce que les patients et les travailleurs de la santé adaptent leur comportement à la procédure d'assignation s'ils préfèrent une intervention à une autre
Étude comparative avant- après	<ul style="list-style-type: none"> • Représente parfois la seule solution pratique, p. ex., pour des interventions à grande échelle où l'assignation aléatoire est impossible pour des raisons pratiques ou politiques 	<ul style="list-style-type: none"> • Des différences connues ou inconnues entre les groupes comparés peuvent exercer une plus grande influence sur les conclusions que le fait que les groupes aient fait l'objet d'interventions différentes. En conséquence, il pourrait être risqué d'établir des relations de cause à effet • Exige de disposer de données de référence
Étude de série chronologique interrompue	<ul style="list-style-type: none"> • Peut être réalisable et relativement facile à mener si l'on dispose des données nécessaires. Aucun groupe témoin requis 	<ul style="list-style-type: none"> • L'ampleur de l'effet est toujours difficile à estimer dans ce genre d'analyses parce que des influences autres que l'intervention à l'étude peuvent influencer sur les changements observés
Étude comparative avec groupe témoin historique	<ul style="list-style-type: none"> • Peut être réalisée rapidement et facilement si les données requises sont disponibles 	<ul style="list-style-type: none"> • Des différences connues ou inconnues entre les groupes comparés peuvent exercer une plus grande influence sur les conclusions que le fait que les groupes aient fait l'objet d'interventions différentes. En conséquence, il pourrait être risqué d'établir des relations de cause à effet

	Forces	Faiblesses
Étude de cohorte	<ul style="list-style-type: none"> • Il s'agit souvent d'études de grande envergure comportant un degré élevé de validité externe (c'est-à-dire que les conclusions peuvent être généralisées). Elles s'étendent souvent sur plusieurs années, ce qui permet de déterminer les effets à long terme d'une intervention 	<ul style="list-style-type: none"> • Les études de cohorte sont habituellement longues et coûteuses, surtout en raison de la nécessité de suivre un nombre (généralement) élevé de participants • Des différences connues ou inconnues entre les groupes comparés peuvent exercer une plus grande influence sur les conclusions que le fait que les groupes aient fait l'objet d'interventions différentes. En conséquence, il pourrait être risqué d'établir des relations de cause à effet
Étude cas-témoin	<ul style="list-style-type: none"> • Peut être menée plus rapidement et facilement qu'une étude de cohorte 	<ul style="list-style-type: none"> • La nature rétrospective de telles études implique la collecte de données sur des événements passés. De tels écarts dans le temps peuvent occasionner des erreurs • Des différences connues ou inconnues entre les groupes comparés peuvent exercer une plus grande influence sur les conclusions que le fait que les groupes aient fait l'objet d'interventions différentes. En conséquence, il pourrait être risqué d'établir des relations de cause à effet

	Forces	Faiblesses
Étude transversale	<ul style="list-style-type: none"> • Ne nécessite pas de suivi et peut donc être menée rapidement et souvent à peu de frais 	<ul style="list-style-type: none"> • Des différences connues ou inconnues entre les groupes comparés peuvent exercer une plus grande influence sur les conclusions que le fait que les groupes aient fait l'objet d'interventions différentes. En conséquence, il pourrait être risqué d'établir des relations de cause à effet
Étude qualitative	<ul style="list-style-type: none"> • Permet de recueillir de l'information plus approfondie que d'autres méthodes quantitatives. Permet de comprendre comment les interventions et les programmes fonctionnent (ou ne fonctionnent pas) 	<ul style="list-style-type: none"> • Ne génère pas de données utilisables pour estimer l'effet d'une intervention au-delà de la perception des personnes ayant fait l'objet d'une entrevue ou d'une enquête