

Réduire le déficit des connaissances dans la prestation de soins de santé à l'échelle mondiale

Apports et limites des expérimentations aléatoires

*Andres GARCHITORENA, Megan B. MURRAY, Bethany HEDT-GAUTHIER,
Paul E. FARMER et Matthew H. BONDS*

Contexte : les expérimentations aléatoires dans le domaine de la médecine et de la santé mondiale

La recherche d'une évaluation empirique, systématique et rigoureuse de l'efficacité des interventions destinées à la population humaine n'a jamais cessé d'être une préoccupation pour la médecine, bien avant toute autre discipline. L'évaluation par assignation aléatoire (*Randomized Controlled Trials – RCT*) est parmi ces démarches d'évaluation rigoureuses, dans laquelle les chercheurs affectent des statuts de traitement à des individus sélectionnés de manière aléatoire avant de comparer les résultats. Soutenus par une industrie pharmaceutique florissante, les essais contrôlés ont été progressivement adoptés au cours des XVIII^e et XIX^e siècles (BOTHWELL et PODOLSKY, 2016). Ceux-ci visaient à distinguer les produits médicaux efficaces (comme les vaccins, les antibiotiques) des nombreux remèdes, thérapies ou répliques aux effets positifs douteux (BOTHWELL et PODOLSKY, 2016). Au cours de la première moitié du XX^e siècle, les chercheurs

ont souvent eu recours à d'autres modèles d'affectation (en traitant par exemple un patient sur deux), mais ceux-ci entraînaient d'importants biais de sélection, les médecins sélectionnant les patients en fonction des besoins qu'ils percevaient. L'épidémiologiste Austin Bradford Hill s'est attaqué à cette problématique en 1948 en commençant une série d'essais sur le traitement de la tuberculose qui mettait en œuvre une randomisation aveugle stricte des patients (autrement dit, des « évaluations randomisées »). Soutenues par le British Medical Research Council et rapidement adoptées par la communauté des chercheurs, les RCT se sont vite hissées au premier rang des modèles expérimentaux en recherche clinique. En 1970, la Food and Drug Administration américaine a exigé de l'industrie pharmaceutique qu'elle fournisse des résultats de RCT avant d'autoriser la mise sur le marché de tout nouveau médicament, conférant alors un rôle central aux RCT dans les réglementations et directives internationales (BOTHWELL et PODOLSKY, 2016).

L'idée de promouvoir les RCT au rang d'étalon-or de la recherche clinique a été favorisée par un mouvement en faveur de la médecine basée sur les preuves (en anglais *evidence-based medicine* – EBM), qui vise à améliorer la pratique clinique par une évaluation critique de la littérature scientifique pour permettre aux cliniciens d'adopter les meilleures pratiques. Largement influencé par le livre d'Archie Cochrane intitulé *Effectiveness and Efficiency: Random Reflections on Health Services*, paru en 1972, la médecine basée sur les preuves s'appuie sur le classement hiérarchique de la qualité des études d'efficacité en fonction de la méthodologie utilisée, les RCT figurant en tête de ce classement et les études observationnelles sans groupe de contrôle en queue de peloton. Au cours des décennies suivantes, les RCT ont gagné en popularité hors du cadre de la recherche clinique. Dans les pays occidentaux, leur utilisation s'est étendue à l'évaluation des politiques publiques dans les domaines de l'éducation, de l'économie, de la sociologie et de la santé publique. Le développement des essais randomisés par grappes à la fin des années 1970, qui randomisaient des groupes de sujets plutôt que des individus isolés, a permis une application encore plus large aux évaluations pour lesquelles la randomisation individuelle était irréalisable ou indésirable. L'utilisation des RCT dans les domaines de la santé mondiale et du développement international a cependant pris du retard, avec seulement quelques dizaines d'études publiées avant les années 2000 (CAMERON *et al.*, 2016).

Un tournant majeur a été pris au début du XXI^e siècle avec l'établissement des Objectifs du millénaire pour le développement (OMD). Reconnaisant que la santé est à la fois un objectif central du développement et un moteur potentiel (SACHS, 2001), les Nations unies ont veillé à ce que les résultats en matière de santé y figurent en bonne place, avec des engagements à réduire la mortalité infantile (OMD 4), à améliorer la santé maternelle (OMD 5) et à combattre le sida, le paludisme et d'autres maladies (OMD 6). Les financements alloués à ces domaines sont montés en flèche (Institute for Health Metrics and Evaluation [IHME], 2016), les fondations privées ont commencé à jouer un rôle de plus en plus important et de grandes organisations comme GAVI, l'Alliance du vaccin, et le Fonds mondial de lutte contre le sida, la tuberculose et le paludisme ont

été créées pour canaliser ces efforts internationaux. Il est alors devenu urgent de mesurer rigoureusement l'impact des interventions sur l'atteinte des objectifs et de disposer d'informations en vue de leur mise à l'échelle. L'intérêt porté à une approche de la santé mondiale fondée sur des preuves empiriques s'est ainsi accru, à l'instar de la révolution de la pratique clinique qui avait eu lieu quelques décennies plus tôt. Selon CAMERON *et al.* (2016), 92,8 % de toutes les évaluations d'impact en matière de santé menées de 2000 à 2012, qui sont répertoriées dans le référentiel des évaluations d'impact (International Initiative for Impact Evaluation), étaient des RCT, la moyenne dans d'autres domaines du développement international étant de 66,4 % (CAMERON *et al.*, 2016).

À la lumière de l'importance prise par les RCT dans le domaine de la santé mondiale, nous étudions ci-après les apports principaux des évaluations randomisées en la matière et mettons en évidence les limites qui pourraient être traitées de façon plus appropriée avec d'autres méthodes de recherche dans le cadre des politiques et des pratiques de santé mondiale.

Les apports en matière de politiques et de pratiques

Depuis l'an 2000, des centaines de RCT ont servi de base à l'élaboration de directives internationales dans des domaines prioritaires de la santé mondiale. Pour atteindre les OMD 4 et 5, les travaux d'évaluation en matière de santé maternelle et infantile se sont focalisés sur la réduction des retards de croissance des fœtus et des enfants, du dépérissement et des carences en micronutriments, qui comptent parmi les facteurs majeurs de mortalité infantile (BHUTTA *et al.*, 2013). Les données issues de ces essais ont par exemple servi de base à la série d'actions recommandées par l'OMS (Organisation mondiale de la santé) et l'Unicef (United Nations International Children's Emergency Fund) pour la supplémentation en fer, en acide folique et en calcium des femmes enceintes, la promotion de l'allaitement maternel et l'apport de vitamine A et de zinc aux enfants (BHUTTA *et al.*, 2013). Les RCT ont également fourni des données empiriques sur le développement des plateformes de prestation communautaire de soins visant à traiter les maladies infantiles courantes telles que la diarrhée, le paludisme et les infections respiratoires, et démontré que de telles approches peuvent réduire les taux de mortalité infantile (HATT *et al.*, 2015 ; WHIDDEN *et al.*, 2018). Bien que les OMD 4 et 5 n'aient pas été atteints, la mise à l'échelle des interventions évaluées par le biais des RCT au niveau des plateformes de soins primaires et de prestation communautaire a contribué à réduire le taux de mortalité des enfants de moins de cinq ans de 53 % (soit un passage de 90,6 à 42,5 décès pour 1 000 naissances vivantes) (YOU *et al.*, 2015), et le taux de mortalité maternelle de 43,9 % (de 385 à 216 décès pour 100 000 naissances vivantes) (ALKEMA *et al.*, 2016) entre 1990 et 2015.

La majorité des RCT liées à la santé, réalisées entre 2000 et 2013 dans les pays à revenus faibles et intermédiaires, consistaient essentiellement à évaluer des interventions biomédicales isolées pour la prévention, le diagnostic ou le traitement d'une maladie particulière (CAMERON *et al.*, 2016 ; KELAHER *et al.*, 2016). Plus de 1 300 RCT ont ainsi été menées sur le VIH (virus d'immunodéficience humaine ou sida) (763 RCT), le paludisme (665 RCT) et la tuberculose (165 RCT) (KELAHER *et al.*, 2016). Elles ont fourni des preuves solides en faveur d'interventions directes de prévention du VIH, comme la prophylaxie antirétrovirale avant exposition, la circoncision médicale volontaire des hommes, ainsi que de nouveaux traitements efficaces (KRISHNARATNE *et al.*, 2016). En outre, les RCT relatives au paludisme ont permis de tester de nouveaux traitements pour les cas de paludisme simples, tels que les polythérapies à base d'artémisinine, ainsi que des approches préventives comme les moustiquaires imprégnées d'insecticide, le traitement préventif intermittent du paludisme pendant la grossesse ou la prophylaxie du paludisme chez les enfants (BHUTTA *et al.*, 2013 ; MARTINEZ-ALONSO et RAMOS, 2016). Les interventions communautaires contre la tuberculose associées à une thérapie sous observation directe se sont avérées très efficaces pour améliorer les taux d'observance et de réussite du traitement (ARSHAD *et al.*, 2014 ; South African Cochrane Centre, 2014), tandis que l'utilisation de nouveaux traitements préventifs s'est révélée efficace pour la prévention de la tuberculose chez les personnes infectées ou non par le VIH (South African Cochrane Centre, 2014). Les programmes actuels d'administration de masse de médicaments, qui constituent le pilier des stratégies de contrôle et d'élimination de nombreuses maladies tropicales négligées (par exemple les géohelminthiases transmises par le contact avec le sol, la filariose lymphatique ou la schistosomiase), ont été étendus à l'échelle internationale suite à des RCT (KAPPAGODA et IOANNIDIS, 2014). Parmi les exemples notables, on peut citer une RCT réalisée en 2004 par Kremer et Miguel, qui a montré les effets significatifs des médicaments antiparasitaires sur l'absentéisme et les performances scolaires au Kenya (MIGUEL et KREMER, 2004), et a conduit dans le monde entier à la mise en œuvre de programmes de déparasitage de portée nationale (HATT *et al.*, 2014). Il ne s'agit là que de quelques exemples où les RCT ont permis à la communauté sanitaire mondiale de disposer d'un ensemble d'interventions pour réduire l'incidence des maladies. La mise à l'échelle de ces interventions (parmi de nombreux autres facteurs) a contribué à réduire la mortalité due au paludisme de 58 % et les nouvelles infections par le VIH de 40 % entre 1990 et 2015. On estime par ailleurs que 37 millions de décès dus à la tuberculose ont pu être évités au cours de la même période (United Nations, 2015).

Un certain nombre de RCT importantes ont également fourni des informations sur des politiques et réformes de santé publique plus larges avant leur mise en œuvre à l'échelle nationale ou internationale (GERTLER, 2004). Un exemple classique est l'évaluation du programme mexicain Progresa, consistant en transferts monétaires pour permettre aux ménages pauvres de participer à une série d'activités liées à la santé comme les soins prénataux, la protection infantile, la vaccination, le suivi nutritionnel, ainsi que des programmes éducatifs de

promotion de la santé. Les essais, qui ont alloué ces aides de manière aléatoire à des groupes spécifiques, ont montré une réduction cohérente des taux de maladie dans les groupes ayant bénéficié de l'action par rapport à la population de contrôle (GERTLER, 2004). Des programmes de transferts monétaires ont depuis lors été testés et mis en œuvre dans le monde entier (HATT *et al.*, 2014). Les RCT ont également apporté des éclairages sur des réformes des soins de santé, comme les programmes de financement basé sur la performance (FBP). Une évaluation initialement randomisée de cette stratégie de FBP appliquée au Rwanda a démontré son impact, et fourni des données pour le déploiement national qui a été mené par la suite dans le cadre de la refonte du système de santé rwandais (KRUK *et al.*, 2016). Depuis, plus de 20 pays africains ont lancé ou commencé à intensifier des programmes FBP dans le domaine des soins de santé (MEESSEN *et al.*, 2011). Les résultats d'une série de RCT ont également nourri le débat sur l'introduction de copaiements d'un montant peu élevé pour des services préventifs et curatifs. Il est généralement admis que les copaiements sont essentiels pour promouvoir la durabilité, réduire les abus et garantir une utilisation judicieuse des produits et services (BATES *et al.*, 2012). Les RCT ne cessent toutefois de démontrer que le fait de facturer des frais peu élevés pour des produits préventifs tels que le savon, les moustiquaires, les vermifuges ou les agents de désinfection de l'eau réduit considérablement l'accès à ces produits pour ceux qui en ont le plus besoin, alors que cela ne génère que peu de recettes (BATES *et al.*, 2012). Bien que l'adoption généralisée de ces idées ait pris du retard, les gouvernements proposent aujourd'hui la plupart de ces produits à titre gratuit dans le cadre de leurs politiques nationales de santé.

Malgré le très grand nombre de preuves empiriques produites par les RCT dans des domaines clés de la santé mondiale, les approches sectorielles (qui présentent des avantages transversaux, mais sont intrinsèquement plus complexes) se prêtent moins aux RCT (FRIEDEN, 2017 ; DEATON et CARTWRIGHT, 2018). Le fait de considérer la RCT comme un étalon-or universel en matière de méthodes d'évaluation d'impact et comme une condition préalable à l'élargissement des interventions peut avoir des conséquences inattendues sur les politiques de santé.

Des conséquences inattendues : un écart croissant en matière de preuves empiriques et de financement dans des secteurs clés de la santé

Jusqu'à l'établissement des OMD, il existait un équilibre relatif – et de nombreux débats – sur l'efficacité comparée des interventions sanitaires verticales et des interventions horizontales plus intégrées (visant des

« systèmes »). L'ère des OMD a rompu cet équilibre, la plupart des financements et des efforts étant concentrés sur un éventail de programmes verticaux. Du fait de l'urgence et de l'attention portée à quelques domaines prioritaires, les programmes verticaux ont été privilégiés, car ils étaient supposés permettre une plus grande spécialisation des services, accorder davantage de place aux maladies hautement prioritaires, accroître la responsabilisation, produire des résultats plus rapides et avoir de meilleures chances de réussite dans les États faibles (ATUN *et al.*, 2008). L'aide au développement en matière de santé a connu une croissance exponentielle au profit des programmes verticaux visant la santé des enfants (par exemple, la vaccination, la malnutrition), la santé maternelle, le VIH, le paludisme et la tuberculose, passant d'environ 3 à 4 milliards de dollars par an en 1990-2000 à plus de 24 milliards en 2016 (Institute for Health Metrics and Evaluation [IHME], 2016). Cette évolution a été soutenue en parallèle par l'augmentation des preuves empiriques issues des RCT pour la mise à l'échelle d'actions efficaces (voir section précédente).

Malgré leurs avantages, les programmes verticaux sont des approches descendantes pilotées de l'extérieur qui, en l'absence d'investissements parallèles dans des systèmes de santé plus solides, peuvent avoir des répercussions négatives comme la fragmentation des services, l'augmentation des obstacles à l'accès aux soins de santé pour les populations non ciblées et la réduction de l'efficacité et de la durabilité des systèmes de santé (ATUN *et al.*, 2008). Les interventions horizontales, comme le renforcement des systèmes de santé et les approches sectorielles, sont complexes par nature, nécessitent une adaptation spécifique au contexte et agissent à de multiples niveaux d'un système de santé (PLSEK et GREENHALGH, 2001 ; CAMPBELL *et al.*, 2007). Leur évaluation au moyen de RCT présente des difficultés importantes, nécessite des investissements conséquents, et s'avère bien souvent irréalisable (PLSEK et GREENHALGH, 2001 ; CAMPBELL *et al.*, 2007). Comme les actions de renforcement des systèmes de santé et les approches sectorielles souffrent à la fois d'un faible engagement politique et d'un manque de preuves d'efficacité fondées sur des RCT, le pourcentage de l'aide au développement pour la santé qui leur est alloué a diminué, passant d'environ 15 % en 1990 à moins de 10 % en 2016 (fig. 1) (Institute for Health Metrics and Evaluation [IHME], 2016). Comparées à la plupart des domaines d'intervention liés à la santé, les augmentations des financements au profit des actions de renforcement des systèmes de santé et des approches sectorielles n'ont pas suivi la tendance dominante. Les hausses annuelles moyennes des financements dans ce domaine sont passées de 11,4 % pour la période 1990-1999 à 7,1 % pour la période 2000-2009 (alors que tous les autres domaines enregistraient une hausse de ces taux de croissance). Avec une baisse en valeur absolue de 2,3 % subie entre 2010 et 2016, c'est l'un des seuls domaines d'action de la santé à avoir connu une telle évolution au cours de cette période (IHME, 2016).

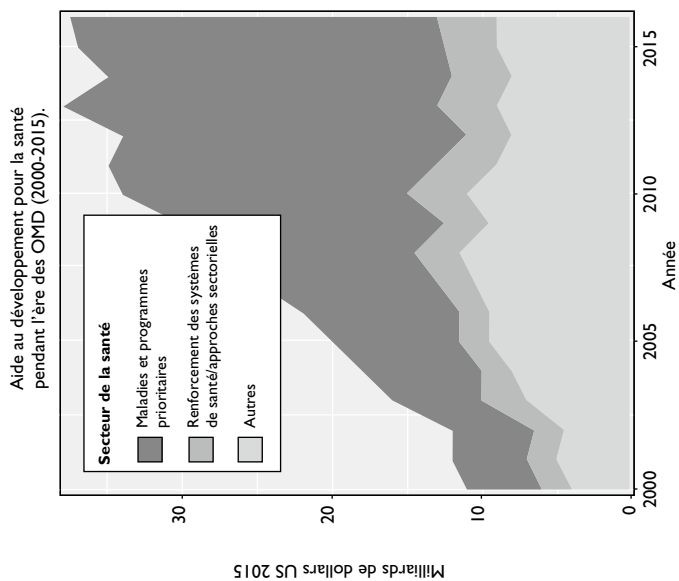
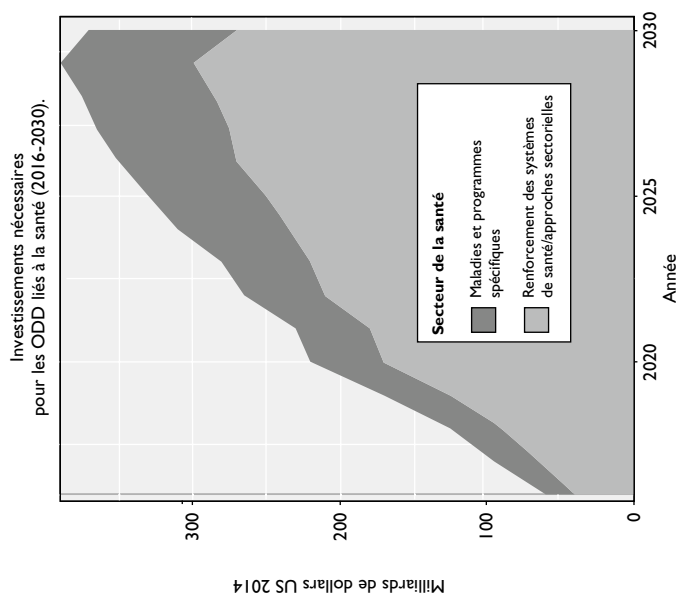


Figure 1
Évolution des investissements en matière de santé entre l'ère des Objectifs du millénaire pour le développement (OMD) (2000-2015) et l'ère des Objectifs de développement durable (ODD) (2016-2030).

Source : Andres Garchitorena, Megan Murray, Bethany Hedt-Gauthier, Paul Farmer et Matthew Bonds.

Note : la figure de gauche représente l'aide totale au développement allouée à la santé par an, adaptée des données de l'IHME (2016). La figure de droite, adaptée des données de STEMBERG *et al.* (2017), représente la projection des investissements supplémentaires nécessaires dans 67 pays à revenus faibles et intermédiaires pour atteindre l'Objectif de développement durable (ODD) 3 lié à la santé. Elle montre qu'une augmentation considérable des financements pour le renforcement des systèmes de santé (RSS) et les approches sectorielles est nécessaire au cours des quinze prochaines années.

Le programme post-OMD des Nations unies, qui s'articule autour des Objectifs de développement durable (ODD), traduit une tentative pour réduire cet écart en se concentrant explicitement sur des approches sectorielles comme la couverture sanitaire universelle (CSU) et le renforcement des systèmes de santé. L'OMS a estimé que pour atteindre les ODD liés à la santé, près des trois quarts de l'ensemble des investissements supplémentaires requis pour les pays à revenus faibles et intermédiaires devraient être alloués aux mesures de renforcement des systèmes de santé et aux approches sectorielles au cours de la période 2015-2030, ce qui représente environ 300 milliards de dollars par an d'ici 2030 (STENBERG *et al.*, 2017) (fig. 1). Une évolution aussi radicale nécessite de repenser en profondeur les données nécessaires et les méthodes d'évaluation appropriées pour éclairer les décisions sur l'affectation et la mise en œuvre des financements. Les organisations internationales comme la Banque mondiale, l'OMS et l'United States Agency for International Development (USAID), entre autres, s'accordent de plus en plus à dire que la base de connaissances actuelle permettant d'orienter cette intégration horizontale est terriblement inadaptée, malgré ses avantages apparents (ATUN *et al.*, 2008 ; GIEDION *et al.*, 2013 ; HATT *et al.*, 2015). Une revue des études systématiques sur le renforcement des systèmes de santé réalisée par l'USAID en 2015 a notamment conclu qu'« il est nécessaire de trouver des méthodes complémentaires pour estimer les effets des mesures de renforcement visant des systèmes adaptatifs complexes » (HATT *et al.*, 2015).

Le manque de preuves empiriques dans des domaines clés de la santé tels que le renforcement des systèmes de santé et la couverture maladie universelle est symptomatique d'un problème plus large de la recherche dans le domaine de la santé : le décalage entre la portée des questions traitées par les RCT et le type de données nécessaires pour améliorer les résultats en matière de santé. On estime que 97 % des fonds de recherche sont consacrés au développement de nouvelles technologies de santé (essentiellement des produits pharmaceutiques), alors que 3 % seulement vont à la recherche sur la mise en œuvre (KRUK *et al.*, 2016). L'efficacité de la configuration des programmes et de l'optimisation de leur exécution est de ce fait rarement évaluée, ce qui entraîne des lacunes importantes pour la transposition des interventions en conditions réelles (KRUK *et al.*, 2016).

Défis et limites

L'une des questions les plus importantes en matière de santé mondiale est de savoir pourquoi des technologies connues – celles qui ont fait leurs preuves dans certains contextes – échouent systématiquement à toucher les personnes auxquelles elles sont destinées. La moitié de la population mondiale n'a pas accès à des services de santé essentiels (World Health Organization et World Bank, 2017). La majorité des décès d'enfants en Afrique subsaharienne est due à des maladies – diarrhée, paludisme, pneumonie – contre lesquelles les solutions

sont connues, peu coûteuses et efficaces. La thérapie de réhydratation orale, par exemple, permet d'éviter 90 % des décès d'enfants liés à la diarrhée dans le monde, mais seuls 4 enfants sur 10 qui en ont besoin en bénéficient (KRUK *et al.*, 2016). Dans la majorité des pays en développement, les ministères de la Santé ont établi des politiques nationales basées sur des normes internationales, mais, dans bien des cas, on ne sait pas quelle est la meilleure façon de les mettre en œuvre, même à petite échelle. Le défi est que même des technologies simples nécessitent des systèmes de prestation de soins complexes – personnels de santé formés, infrastructures, fournitures et médicaments – pour leur mise en œuvre sur le lieu de prise en charge. Les défaillances se manifestent à différents niveaux, depuis les agents de santé communautaires individuels aux établissements de soins de santé, en passant par les chaînes d'approvisionnement nationales, et se rétro-alimentent (BRUMMITT *et al.*, 2017). Ce phénomène a motivé le mouvement en faveur du « renforcement des systèmes de santé ». L'utilisation des RCT pour répondre à ces questions fondamentales comporte des difficultés intrinsèques. Pour un examen approfondi des limites méthodologiques des RCT, se reporter aux chap. 1 (Ravallion) et 2 (Pritchett) de ce volume, ou au débat suscité récemment par DEATON et CARTWRIGHT (2018). Nous présentons ci-dessous des sujets choisis relatifs à la pratique de la santé mondiale.

Le point fort d'une RCT bien menée est sa solide validité interne. Dans des conditions contrôlées, elle peut fournir une estimation non biaisée de l'effet de traitement moyen pour des interventions particulières, bien que les critiques affirment que cela n'est vrai que si de nombreuses hypothèses concernant la conception de l'essai sont confirmées (COOK, 2018). Même lorsque les résultats des RCT produisent une estimation non biaisée de l'effet pour la population spécifique étudiée, cela ne présage pas nécessairement l'effet de l'action sur d'autres populations en conditions réelles, ce qui signifie que l'on observe fréquemment des différences entre l'impact d'une action mise en œuvre dans des conditions contrôlées (« *efficacy* ») et son effet véritable dans le monde réel (« *effectiveness* ») (AHMED *et al.*, 2010 ; SHELTON, 2014). Contrairement aux études observationnelles, le processus de recrutement des participants à une RCT peut remplacer artificiellement les systèmes de prestation en place dans le monde réel afin de créer les conditions d'étude optimales, ce qui est particulièrement inutile car l'adhésion à un système de prestation performant constitue elle-même un problème central à résoudre. En outre, l'impact des interventions, tel que mesuré par les RCT réalisées dans des populations ou contextes différents, peut varier considérablement. Dans l'exemple du déparasitage cité plus haut, par exemple, les méta-analyses de plusieurs RCT ne permettent pas d'établir de manière concluante si les campagnes de masse ont un effet sur l'état nutritionnel, les résultats scolaires ou la survie des enfants, étant donné l'hétérogénéité des résultats des études et les différences dans les critères d'inclusion dans chaque méta-analyse (TAYLOR-ROBINSON *et al.*, 2015 ; CROKE *et al.*, 2016 ; VRIEZE, 2018).

Le coût élevé des RCT peut conduire les chercheurs à prendre en compte des périodes d'étude ou des tailles d'échantillons insuffisantes pour évaluer

correctement l'effet du traitement, ou à utiliser des indicateurs indirects qui ne correspondent pas au résultat évalué ou qui ne permettent pas de le déterminer (par exemple des indicateurs de processus, des signes limités ou précoces de maladie) (FRIEDEN, 2017). À titre d'exemple, la suppression des paiements directs au point d'intervention par le biais des régimes d'assurance ou de dispenses de frais pour le bénéficiaire des services de soins constitue une stratégie clé pour accroître l'accès aux soins de santé, fournir une couverture financière contre des dépenses catastrophiques et améliorer finalement les résultats en matière de santé. Alors que la suppression des frais pour le bénéficiaire des soins peut impacter certains résultats sanitaires et économiques dans des populations de tous âges, une RCT réalisée au Ghana n'a mesuré que l'anémie entraînée par le paludisme chez les enfants de moins de cinq ans. L'essai a conclu que cette action n'avait pas d'effets mesurables sur la santé (ANSAH *et al.*, 2009), mais l'évaluation a été effectuée six mois seulement après le début de l'intervention, et l'étude était sous-dimensionnée compte tenu de la faible prévalence de l'anémie observée (RIDDE et HADDAD, 2009). Les problèmes de conception tels que ceux illustrés dans cet exemple, bien qu'ils ne soient pas l'apanage des seules RCT, sont relativement courants. Une revue des RCT indexées dans PubMed publiées en 2001 et 2006 a révélé que de nombreuses évaluations étaient menées avec des tailles d'échantillons insuffisantes pour détecter même des effets de traitement importants (HOPEWELL *et al.*, 2010). Des conceptions d'études inadéquates, entre autres facteurs, contribuent à un gaspillage important des investissements dans la recherche biomédicale, estimé à environ 85 % des ressources investies (200 milliards de dollars en 2010) (MACLEOD *et al.*, 2014).

Certains chercheurs privilégient les RCT parce que la randomisation permet d'équilibrer les facteurs connus et inconnus qui influencent le résultat évalué, ce qui simplifie l'inférence statistique dans l'évaluation avec une connaissance minimale des mécanismes qui sous-tendent les effets observés. Ce modèle peut toutefois avoir un effet pervers car il réduit le flux d'informations entre les chercheurs et le contexte dans lequel ils réalisent les essais, facteur particulièrement important lorsque ce contexte implique des populations vulnérables agissant dans des conditions complexes. On peut également aboutir à des conceptions d'études non éthiques (DEATON et CARTWRIGHT, 2018). Les essais expérimentaux éthiques exigent par exemple que l'intervention soit réalisée selon le principe d'équipoise – c'est-à-dire avec une incertitude quant à ses bénéfices –, mais de nombreuses RCT sont effectuées dans le but de confirmer les résultats d'études observationnelles, en évitant d'offrir les avantages du programme aux sujets présents dans le groupe de contrôle (FARMER *et al.*, 2013) (pour plus de détails sur l'équipoise et ses implications pour les RCT, voir Abramowicz et Szafarz, chap. 10, ce volume). Quelle que soit la conception de l'étude, la création d'un ensemble de preuves empiriques dans le domaine des prestations de services en santé mondiale requiert l'existence de larges boucles d'information entretenues sur le long terme, dans lesquelles les acteurs locaux, les praticiens et les exécutants participent activement à la hiérarchisation des problématiques de recherche, ainsi qu'à l'interprétation et à la diffusion des résultats, générant

des possibilités de formation et de recherche encadrées qui apportent des éclairages sur les services fournis (FARMER *et al.*, 2013). Dans la section suivante, nous illustrons la façon dont des cadres d'évaluation complémentaires peuvent contribuer à la constitution de cet ensemble de preuves.

Au-delà des RCT pour les ODD : des cadres d'évaluation observationnelle pour le renforcement des systèmes de santé et la CSU

La recherche sur la mise en œuvre a pour but d'établir si, comment, quand et pourquoi une intervention fonctionne, et de proposer d'autres hypothèses ultérieures (BHATTACHARYYA *et al.*, 2009 ; KRUK *et al.*, 2016). Elle fait appel à divers modèles d'études, allant des méthodes observationnelles et expérimentales quantitatives à la recherche qualitative (KRUK *et al.*, 2016). L'utilisation des résultats d'études observationnelles, qui viennent parfois contredire leurs homologues expérimentales, fait l'objet de nombreux débats (IOANNIDIS *et al.*, 2001 ; PRASAD *et al.*, 2013 ; HEMKENS *et al.*, 2016 ; JONES et STEEL, 2018). De multiples études comparant les essais randomisés et non randomisés montrent cependant que les études observationnelles de grande qualité (comme les études prospectives contrôlées) peuvent produire des résultats comparables à ceux des RCT (IOANNIDIS *et al.*, 2001 ; JONES et STEEL, 2018). Le principal défi des études observationnelles est le risque accru de choisir des groupes de comparaison inappropriés avec des facteurs non mesurés qui peuvent fausser les résultats. Une évaluation initiale du projet « Villages du millénaire » a notamment été critiquée pour avoir choisi rétroactivement un contrôle biaisé qui avait orienté favorablement les résultats de l'étude (MITCHELL *et al.*, 2018). Dans ce cas, les systèmes de données et d'évaluation étaient initialement insuffisants en l'absence de groupes de contrôle *a priori*, ce qui a compromis la capacité à tirer des conclusions définitives sur l'efficacité de l'intervention. Les analyses rétrospectives de suivi se sont toutefois avérées plus solides (MITCHELL *et al.*, 2018). On note également que, dans certains cas, les méthodes observationnelles ont été préférées aux résultats des RCT pour la prise de décision, malgré leur divergence, parce qu'elles permettent des périodes de suivi plus longues, des tailles d'échantillon plus importantes et une probabilité plus forte de détecter des effets négatifs (FRIEDEN, 2017). Les recommandations concernant la vaccination antigrippale par pulvérisation nasale avec des vaccins vivants atténués, qui avaient initialement démontré une bonne protection dans des RCT, ont par exemple évolué dans le temps après que des études observationnelles ultérieures ont suggéré que la validité externe des résultats des RCT était limitée (FRIEDEN, 2017).

Concernant les approches sectorielles, le Medical Research Council britannique reconnaît que la conception, la description et la mise en œuvre d'une intervention complexe sont les points faibles les plus fréquents des RCT (CAMPBELL *et al.*, 2000). Il ne propose pas de méthodologies alternatives, mais fournit des lignes directrices explicites permettant la réalisation de RCT bien conçues pour des interventions complexes (CAMPBELL *et al.*, 2000). Il existe des exemples convaincants d'interventions complexes qui ont été évaluées par le biais de RCT (BANERJEE, *et al.*, 2015a), comme celles décrites ci-dessus pour le Mexique et le Rwanda. Les ressources nécessaires pour mener de telles évaluations limitent toutefois leur capacité à être utilisées à grande échelle dans les pays à revenus faibles et intermédiaires. Parallèlement, les chercheurs tentent de tirer des enseignements issus du domaine des « sciences de la complexité » pour mieux comprendre les systèmes de soins de santé, qui sont à la fois complexes et adaptatifs (PLSEK et GREENHALGH, 2001). La théorie de la complexité suggère qu'au lieu de décomposer le système en éléments simples (comme avec les RCT portant sur de multiples interventions verticales), il peut être préférable de mettre en œuvre simultanément diverses approches, et de s'orienter progressivement vers celle(s) qui fonctionne(nt) (équivalent à une mise en œuvre adaptative avec des études observationnelles quasi expérimentales) (PLSEK et GREENHALGH, 2001). Pour atteindre les objectifs de développement liés à la santé, comme la santé maternelle et infantile, le facteur le plus important peut être l'effet collectif d'une série optimale d'actions destinées à des populations et à des contextes particuliers (SHELTON, 2014). Dans cette optique, cumuler, en parallèle à la myriade d'interventions de prestation de soins de santé qui ont lieu dans les pays en développement, des méthodes solides de collecte de données et d'évaluation (notamment des méthodes observationnelles quasi expérimentales), représente l'une des opportunités majeures en matière de santé mondiale. Cela peut permettre de mener des recherches rigoureuses à moindre coût et sans devoir contrôler le processus de mise en œuvre ou la population bénéficiaire.

Les limites des RCT pour l'évaluation de programmes de santé mondiale complexes déployés à grande échelle ont conduit au développement de cadres qui prennent en compte, parallèlement à la mise en œuvre des programmes, un ensemble de méthodes observationnelles, telle que celles proposés par le Partenariat international pour la santé (IHP+), l'Initiative pour la santé en Afrique et l'Initiative catalytique pour sauver un million de vies (World Health Organization, 2010 ; VICTORA *et al.*, 2011 ; BRYCE *et al.*, 2013), auxquels participent des acteurs majeurs comme l'OMS, la Banque mondiale et la fondation Bill et Melinda Gates. Ces cadres d'évaluation déterminent la réussite des programmes en termes de gains de couverture des interventions et d'effets sur la santé. Les études sont réalisées en conditions réelles avec une mise en œuvre plus variable que dans les essais contrôlés. Les interventions se déroulent rarement de manière isolée, étant donné que de nombreuses organisations mettent en œuvre des programmes pratiquement partout dans les pays en développement, et que des changements s'opèrent au niveau des situations sanitaires et

socio-économiques indépendamment des programmes existants (World Health Organization, 2010 ; VICTORA *et al.*, 2011 ; BRYCE *et al.*, 2013 ; EL-SADR *et al.*, 2014 ; REIDY *et al.*, 2018).

En se servant du district de santé comme unité d'étude, les chercheurs évaluent des indicateurs clés au niveau des intrants, des processus et des extrants du système de santé (par exemple, le personnel de santé, les services disponibles), conjointement à des indicateurs de résultat et d'impact (par exemple, la couverture des services et les taux de mortalité ; tabl. 1). Outre le suivi continu de la mise en œuvre du programme, la collecte de données complémentaires permet aux chercheurs de combler les lacunes d'information avant, pendant et après la période d'évaluation, en utilisant des évaluations des établissements de santé, des enquêtes auprès des ménages, des recherches longitudinales et des études qualitatives. Les analyses quantitatives sont complétées par des descriptions qualitatives de la mise en œuvre des programmes (c'est-à-dire ce qui est mis en œuvre et de quelle manière) et des facteurs contextuels qui peuvent avoir eu une incidence sur la mise en œuvre et l'impact, de façon à ce que les résultats puissent être interprétés de manière appropriée et des enseignements tirés (VICTORA *et al.*, 2011 ; REQUEJO *et al.*, 2015 ; REIDY *et al.*, 2018).

Au niveau national, ces évaluations à grande échelle de l'efficacité des programmes ont conduit à l'initiative « *Countdown* », qui assure le suivi d'une liste complète des indicateurs mentionnés ci-dessus pour chaque pays à revenus faibles ou intermédiaires, et permet ainsi des comparaisons objectives et solides des progrès réalisés par chaque pays (REQUEJO *et al.*, 2015). Au niveau infranational, ce cadre sert à évaluer l'impact des actions complexes de renforcement des systèmes de santé en contribuant à combler le manque important de preuves empiriques en la matière. Prenons à titre d'illustration deux expériences récentes menées au Rwanda (THOMSON *et al.*, 2018) et à Madagascar (GARCHITORENA *et al.*, 2018), qui mettent en œuvre une série similaire d'actions de renforcement du système de santé intégrées à divers niveaux de prestations de soins (santé communautaire, centres de soins de santé primaires, hôpital de référence). Les deux interventions visent à améliorer la capacité du système de santé par le biais de programmes horizontaux tout en intégrant verticalement des programmes cliniques prioritaires. Des évaluations sont effectuées à l'aide des données recueillies au niveau des ménages dans le cadre d'enquêtes démographiques et de santé transversales, et émanant d'échantillons représentatifs des populations respectives (pour Madagascar, la démarche incluait une enquête de base représentative à la fois de la zone de couverture de l'intervention et de la zone de comparaison), et d'enquêtes répétées à intervalles rapprochés pendant toute la durée des interventions (tous les cinq ans pour le Rwanda et tous les deux ans pour Madagascar). L'impact des interventions est évalué au moyen d'analyses statistiques similaires à la méthode des différences de différences pour un large éventail d'indicateurs de résultat du type de ceux présentés dans le tabl. 1, en contrôlant par des facteurs confondants pertinents (par exemple, la richesse des ménages).

Tableau 1
Indicateurs de couverture sanitaire et de mortalité dans le continuum de soins
pour la santé maternelle et infantile.

Indicateurs de couverture sanitaire (%)	Indicateurs de mortalité
Prégrossesse	
Demande de planification familiale satisfaite	
Grossesse	
Suivi prénatal (≥ 1 visite)	
Suivi prénatal (≥ 4 visites)	
Traitements préventifs intermittents du paludisme durant la grossesse	
Protection contre le tétanos néonatal	
Naissance	
Accouchement assisté par du personnel soignant qualifié	Mortalité maternelle (nombre de décès pour 100 000 femmes)
Période postnatale	
Visite postnatale pour les mères	Mortalité néonatale (nombre de décès pour 1 000 naissances vivantes)
Visites postnatales pour les nourrissons	
Initiation précoce à l'allaitement maternel	
Petite enfance	
Allaitement maternel exclusif (< 6 mois)	Mortalité infantile (nombre de décès pour 1 000 naissances vivantes)
Introduction des aliments (6-8 mois)	
Vaccination DTCoq (diphtérie-tétanos-coqueluche), 3 doses	
Première injection du vaccin contre la rougeole	
Vaccination Hib3 (<i>hæmophilus influenzae</i> de type b)	
Supplémentation en vitamine A (2 doses)	
Enfance	
Enfants dormant sous des moustiquaires imprégnées d'insecticide	
Recours aux soins pour les symptômes de la pneumonie	
Traitement antipaludéen de première intention	
Traitement aux sels de réhydratation orale	
Sources d'eau potable améliorées	
Installations sanitaires améliorées	
Indicateurs composites de santé materno-infantile	
Indice de couverture composite (ICC)	

Source : Andres Garchitorena, Megan Murray, Bethany Hedt-Gauthier, Paul Farmer et Matthew Bonds, adaptation à partir des données de REQUEJO et al. (2015).

Ce modèle d'étude quasi expérimental permet aux responsables de programmes de disposer de l'autorité nécessaire sur la mise en œuvre des programmes (quand, où et comment les actions sont déployées) qui n'est pas prescrite par un protocole de recherche. Des systèmes de données sont construits autour de l'intervention de façon à ce que les chercheurs puissent évaluer les activités en cours et fournir des informations susceptibles d'aider les responsables à adapter les programmes sans perturber la mise en œuvre. Une analyse réalisée sur la période 2014-2016

à Madagascar a par exemple montré que, malgré des améliorations globales de la plupart des indicateurs de couverture, l'accès aux soins de santé restait très faible pour les populations éloignées des établissements de santé (GARCHITORENA *et al.*, 2018). Ce constat a entraîné une expansion du soutien aux programmes de santé communautaire, tant géographiquement que dans la portée des services fournis, sans que la conception de l'étude ne s'y oppose. En outre, des études complémentaires de recherche sur la mise en œuvre permettent d'évaluer des composants spécifiques au sein de l'intervention globale, comme un programme de mentorat et de supervision accrue, mené dans le cadre de l'intervention de renforcement du système de santé au Rwanda (MANZI, *et al.*, 2018a ; 2018b), tout en contribuant à améliorer les capacités de recherche au niveau des praticiens locaux (HEDT-GAUTHIER *et al.*, 2017 ; ODHIAMBO *et al.*, 2017).

Suite à l'intervention de renforcement du système de santé dans un district et demi du Rwanda rural, la mortalité des enfants de moins de cinq ans a baissé de plus de 60 % entre 2005 et 2010 (THOMSON *et al.*, 2018). Cette réduction était beaucoup plus importante que celle atteinte dans le reste du pays, et trois fois supérieure au taux requis pour satisfaire les OMD. De même, la mortalité des enfants de moins de cinq ans et la mortalité néonatale ont baissé respectivement de près de 20 % et 35 % au cours des deux premières années de l'intervention de renforcement du système de santé dans un district de Madagascar (2014-2016), à un rythme significativement plus rapide que les taux nationaux moyens observés pour n'importe quel pays pendant la période des OMD. Même si les caractéristiques de référence étaient similaires en termes de revenu par habitant et de taux de mortalité des enfants de moins de cinq ans, chaque intervention s'est déroulée dans des contextes politiques et économiques très différents (BONDS et RICH, 2018). Au cours de la période 2005-2010, le Rwanda a connu un cycle vertueux de stabilité politique, d'investissements internationaux et d'aide étrangère. À Madagascar, en revanche, la situation politique a été instable pendant la majeure partie des 50 dernières années, avec une économie en déclin constant et les investissements dans le système de santé les plus faibles à l'échelle mondiale en 2014. À elles deux, ces expériences constituent un test naturel de la portée des impacts que les interventions intégrées de renforcement des systèmes de santé peuvent avoir au niveau de la population, et qui peuvent être reproduits dans des contextes différents (BONDS et RICH, 2018).

Conclusion

Les deux dernières décennies ont vu une amélioration sans précédent des indicateurs de santé dans les pays à revenus faibles et intermédiaires, largement liée aux OMD. Les RCT ont joué un rôle essentiel au cours de cette période, en facilitant l'adoption de technologies et de services médicaux efficaces pour réduire l'incidence des maladies. La plupart de ces services ont fait l'objet de

tentatives de mise à l'échelle verticale, qui peut elle-même se prêter à des RCT. Pourtant, d'importantes régions du monde continuent à souffrir d'un manque d'accès aux soins de santé primaires. Les échecs rencontrés dans le processus de mise à l'échelle sont souvent dus à des défaillances dans la mise en œuvre de base et à la faiblesse des systèmes de santé. Un consensus se fait ainsi autour de l'importance centrale des approches sectorielles comme le renforcement des systèmes de santé intégrés, les soins primaires intégrés et la couverture sanitaire universelle. Pour atteindre les objectifs de développement durable liés à la santé (ODD 3), les investissements dans ces secteurs devront être multipliés par plus de cinq au cours des 15 prochaines années. Quel type de preuves empiriques devrait guider ces investissements et éclairer la mise en œuvre ?

Les approches sectorielles se prêtent moins bien aux RCT, et les modèles d'études devraient être guidés par les priorités programmatiques. Si les RCT sont toujours fondamentales pour déterminer des solutions isolées et pertinentes dans un large éventail de contextes, la recherche sur la mise en œuvre, qui fait généralement appel à des méthodes à la fois qualitatives et quantitatives sans nécessairement randomiser la mise en œuvre, peut aider les responsables des programmes à comprendre comment des actions dont les effets ont été démontrés peuvent être intégrées efficacement dans les systèmes de prestations de soins de santé. Les méthodes observationnelles et quasi expérimentales, en particulier, sont les plus appropriées lorsque l'échelle de l'intervention rend la randomisation impossible ou difficilement réalisable. Le fait d'adjoindre à la myriade d'interventions de prestations de soins de santé organisées dans les pays en développement une collecte de données solide pourrait aider au progrès d'une recherche rigoureuse à faible coût sans nécessité de contrôler la mise en œuvre ou la population bénéficiaire. Un examen complet des différents types de preuves disponibles pourrait contribuer à orienter les efforts déployés en matière de santé mondiale au cours des prochaines décennies.

Expérimentations aléatoires dans le champ du développement

Une perspective critique

Éditeurs scientifiques

Florent BÉDÉCARRATS, Isabelle GUÉRIN et François ROUBAUD

IRD Éditions

INSTITUT DE RECHERCHE POUR LE DÉVELOPPEMENT

Collection Synthèses

Marseille, 2022

Ce texte a fait l'objet d'une première édition en anglais :

BÉDÉCARRATS F., GUÉRIN I., ROUBAUD F. (eds), 2020, Randomized Control Trials in the Field of Development : a Critical Perspective, Oxford, Oxford University Press, 448 p.

<https://global.oup.com/academic/product/randomized-control-trials-in-the-field-of-development-9780198865360>

Photo 1^{re} et 4^e de couverture

© Ermell – Paul Klee, *Klippen am Meer*, 1931, Musée Lenbachhaus, Munich.

https://upload.wikimedia.org/wikipedia/commons/6/69/Klee_Klippen_am_Meer_1290074.jpg

Coordination éditoriale

IRD/Romain Costa

Préparation éditoriale

Marie-Laure Portal (11)

Mise en page

Desk (53)

Maquette de couverture

IRD/Michelle Saint-Léger

Maquette intérieure

IRD/Pierre Lopez



© IRD, 2022

Certains droits réservés. Il s'agit d'une publication en open access, disponible en ligne et distribuée sous les termes de l'attribution Creative commons, non commerciale, non modifiable 4.0 licence internationale (CC BY-NC-ND 4.0), dont une copie est disponible à cette adresse :

<http://creativecommons.org/licences/by-nc-nd/4.0/>

ISBN papier : 978-2-7099-2947-9

ISSN : 2431-7128

ISBN PDF : 978-2-7099-2948-6

ISBN epub : 978-2-7099-2949-3