

Le principe des essais cliniques peut-il s'appliquer aux politiques de développement ?

Publié: 22 janvier 2018, 23:12 CET

Florent Bédécarrats

Chercheur, spécialiste de l'évaluation de projets et politiques, Agence française de développement (AFD)

François Roubaud

Économiste, statisticien, directeur de recherche à l'IRD et membre de l'UMR DIAL, Institut de recherche pour le développement (IRD)

Isabelle Guérin

Directrice de recherche à l'IRD-Cessma, Institut de recherche pour le développement (IRD)



Ma CSV Field Day 2017 Vietnam. Climate Change, Agriculture and Food Security / Flickr, CC BY-NC-SA

En recourant au principe des essais cliniques, des chercheurs prétendent révolutionner notre compréhension des interventions en matière de développement. Si la méthode des essais cliniques est en apparence très attractive, prétendre l'utiliser pour évaluer toutes sortes d'interventions est à la fois problématique et dangereux. C'est ce que montre notre article récemment publié.

Le principe des essais cliniques consiste à tirer au sort deux groupes au sein d'une population homogène : le premier reçoit une « intervention » (médicament, subvention, crédit, formation, etc.), le second un placebo, une intervention différente ou tout simplement rien ; à l'issue d'une certaine période, les deux groupes sont comparés afin de juger de l'efficacité de l'intervention ou d'en analyser deux modalités distinctes. Depuis le milieu du XX^e siècle, cette méthode est couramment appliquée dans le domaine de la médecine, où elle suscite de nombreux débats. Cette méthode a ensuite été transposée à l'évaluation des politiques publiques dans des domaines de l'éducation, la criminalité, la fiscalité, etc., notamment aux États-Unis dans les années 1960-80.

Depuis quelques années, ces essais cliniques (qu'on désigne couramment par leur acronyme anglais : RCT pour, *Randomized Control Trials*) s'ouvrent à un champ nouveau : celui de l'aide aux pays en développement. Une vaste panoplie d'interventions est ainsi passée au crible de la randomisation, notamment en matière d'éducation (incitations visant à réduire l'absentéisme des enseignants, vermifuges destinés à diminuer l'absence des élèves), de santé (filtres à eau, moustiquaires, formations ou systèmes de primes pour le personnel soignant, consultations gratuites, conseils médicaux par SMS, etc.), de finance (microcrédit, microassurance, épargne, éducation financière), ou encore de « gouvernance ».

Un succès sans précédent

Les RCT sont aujourd'hui labélisées d'étalon or (*gold standard*) de l'évaluation, préférable à toutes les autres méthodes qu'on utilise classiquement pour juger si une politique est efficace ou non. Elles sont présentées par ses adeptes comme une véritable révolution copernicienne (voir par exemple les ouvrages de leurs promoteurs les plus en vue, *Repenser la pauvreté* d'Esther Duflo et Abijit Banerjee, et *More than Good Intentions* de Dean Karlan, ou leurs interventions publiques (voir également celle-ci). La communauté académique et politique tend à leur attribuer en exclusive le qualificatif de « rigoureuses », voire de « scientifiques ».

Certains de ses initiateurs, dont la chercheuse franco-américaine Esther Duflo, bénéficient d'une véritable consécration internationale. Rares sont aujourd'hui les formations académiques qui prétendent « tutoyer l'excellence » qui ne proposent un cursus spécialisé dans ce domaine. Rares également les conférences internationales qui ne programment pas une session dédiée aux RCT avec un succès d'assistance réitéré. Rares enfin, les agences d'aides qui n'aient pas créé de département qui leur soit dédié (voir l'exemple de la Banque mondiale, ou de l'agence de coopération au développement des États-Unis, l'USAID) ou qui n'aient engagé ou financé leurs propres RCT.

Non seulement les RCT tendent à occuper une position de plus en plus dominante, mais elles exercent un effet d'éviction sur les autres approches. C'est très clairement le cas pour la Banque mondiale : au cours de la période 2000-2010, à peine 20 % des évaluations étaient des RCT ; dans les cinq années suivantes, les proportions ont été quasiment inversées. Le réseau international 3IE, spécialisé dans l'évaluation, lui a emboîté le pas.

Une telle systématisation des RCT est-elle scientifiquement légitime et politiquement souhaitable ?

De la théorie à la pratique...

Toute évaluation d'impact (d'un projet, d'une politique, d'un programme) se heurte à un défi récurrent : comment isoler l'impact de cette intervention des changements advenus par ailleurs ? De multiples méthodes existent, mais l'avantage des RCT est en théorie incontestable du fait que la sélection aléatoire de grands échantillons garantit, en principe et en moyenne, que toutes les différences mesurées entre les deux groupes sont dues à l'intervention et à rien d'autre.

Mais les RCT ont en réalité bien du mal à répondre aux questions fondamentales sur le développement, et ce pour trois raisons au moins :

- Leur validité externe est faible, c'est-à-dire qu'elles sont très localisées et ne s'appuient pas sur des échantillons représentatifs de la population dans son ensemble. Leurs résultats sont donc difficilement généralisables : impossible de savoir avec ces méthodes si les résultats obtenus dans une zone du rurale du Maroc, s'appliquent à une autre région marocaine, à la Tunisie voisine ou encore en Bolivie. Cet argument est classique et bien admis par tous. Ceux qui suivent le sont moins.
- Contrairement à ce qui est souvent asséné, leur validité interne aussi pose problème. C'est-à-dire que leur capacité à mesurer l'impact de l'intervention évaluée est imparfaite. Comme l'ont bien montré le prix Nobel d'économie 2016, Angus Deaton et sa collègue épistémologue Nancy Cartwright,, les RCT peinent à arbitrer de manière optimale entre biais (à minimiser) et précision (à maximiser) et sont donc amenées à se focaliser sur les résultats *moyens*, pour l'ensemble de la population considérée. Or les impacts des politiques étudiées sont souvent hétérogènes, et cette hétérogénéité est déterminante en matière de politique publique. Par ailleurs, la mise en œuvre des protocoles d'enquêtes se heurte à de nombreuses difficultés d'ordre pratique et éthique, si bien que la comparaison entre population témoin et population traitée est souvent biaisée.
- Une autre raison, souvent inavouée voire expressément occultée, tient au fait que les essais cliniques, dont le coût est souvent proche du million d'euros, mettent en scène une diversité d'acteurs (populations étudiées, ONG, gouvernements, chercheurs, bailleurs de fonds, etc.) aux intérêts multiples, parfois divergents. Il en résulte un jeu d'acteurs qui influence autant le protocole technique et sa mise en œuvre que l'analyse des résultats, leur publication et leur dissémination. Ce bricolage se fait, là encore, au détriment de la rigueur scientifique. Les intérêts en jeu au sein de ces arènes politiques que constituent les RCT concernent tout aussi bien la réélection de gouvernements (exemple du Mexique concernant l'évaluation d'une politique de subvention aux pauvres), la défense d'un discours dominant sur certains outils de développement (exemple de la microassurance), ou bien leur notoriété, notoriété parfois acquise grâce aux promoteurs des RCT (exemple de la controverse sur les vermifuges). Les intérêts en jeu concernent parfois les exigences de publication des chercheurs...

En définitive, le type d'interventions susceptibles d'être évaluées par les RCT est éminemment restreint, quelques pourcents tout au plus selon l'agence de coopération britannique. Circonscrire le champ des évaluations d'impact aux interventions susceptibles de respecter les canons de la randomisation écarte un grand nombre de projets, mais aussi nombre de dimensions structurelles du développement, tant économiques que politiques, comme la régulation des grandes entreprises, la fiscalité, les échanges internationaux, pour n'en citer que quelques-unes.

Émergence d'un nouveau business model de la recherche

Au regard des sommes et des enjeux politiques, inédits dans le domaine de l'évaluation du développement, les RCT peuvent être appréhendées comme une véritable industrie. Comme toute industrie, le marché des évaluations d'impact est la rencontre d'une offre et d'une demande.

Côté demande, bailleurs de fonds et décideurs sont à la recherche d'indicateurs et de données chiffrées permettant d'attester de l'efficacité de leurs politiques et/ou de prendre des décisions. Dans un contexte de raréfaction de ressources et d'exigence croissante de redevabilité des fonds publics, les RCT sont du pain béni. Les résultats sont d'une simplicité séduisante, proposent des réponses univoques à des questions complexes, par exemple le montant moyen de revenu obtenu grâce au microcrédit, la durée moyenne de scolarisation des élèves traités avec des vermifuges.

Certaines études comparent différentes formes d'intervention en fonction de leurs résultats et de leur coûts, offrant ainsi aux décideurs politiques des outils précieux d'aide à la décision. Derrière cette simplicité de résultats se niche toutefois un appareillage statistique complexe, gage de scientificité, mais difficilement accessible (et donc questionnable) par des néophytes.

Toujours côté demande, l'évolution du monde académique est également favorable à la montée en puissance des RCT. Les écoles dites hétérodoxes, centrées sur l'analyse des structures sociales et des processus de domination, n'ont plus le vent en poupe. Priorité est donnée à la quantification, aux fondements micro de processus macro, et au sein des fondements micro, aux ressorts psychologiques et cognitifs des comportements individuels. La montée en puissance de l'économie comportementale, objet de deux prix Nobels au cours de la dernière décennie (2002 puis 2017) conforte celle des RCT, qui en usent et abusent.

Quant à l'offre, elle est largement façonnée par des entrepreneurs scientifiques d'un genre nouveau qui déploient de multiples stratégies pour « tenir » le marché. L'élan en faveur des RCT a surtout été porté par l'émergence d'une nouvelle génération de chercheurs, avec ses figures emblématiques. Ils sont jeunes, issus du sérail des meilleures universités (pour la plupart américaines).

Ils ont su trouver la formule du carré magique en combinant excellence académique (légitimité scientifique), effort de séduction en direction du public (visibilité médiatique et légitimité citoyenne) et des bailleurs de fonds (demande solvable), investissement massif dans la formation (offre qualifiée), et modèle d'entreprise performant (rentabilité financière) ; toutes ces qualités se renforçant mutuellement.

En multipliant des cours dans les cursus universitaires, mais également en proposant des sessions courtes de formation destinées à un large public, en assurant des enseignements classiques (en présentiel) mais également sous des formes nouvelles (formations en lignes ouvertes à tous, « MOOC » en anglais), les randomisateurs se donnent les moyens d'attirer des ressources jeunes, motivées et hautement qualifiées.

En s'engageant dans une intense activité de communication et de plaider, à l'aide de toute une série de supports de presse ou para-académiques (*policy briefcase*, blogs, forums de vulgarisation, *happenings*, etc.), ils montrent l'image avenante de chercheurs ayant accepté de sortir de leur tour d'ivoire. En affichant une posture modeste au plus proche du terrain, ils incarnent l'engagement, l'empathie et le désintéressement.

Plusieurs stratégies sont mises en œuvre pour asseoir le monopole et éviter un rééquilibrage au tout-RCT. Comme ce fut le cas en médecine, les méthodes alternatives sont disqualifiées, les RCT s'arrogeant le monopole de la scientificité.

La référence à l'*evidence-based medicine* est mobilisée comme gage de scientificité, mais en occultant les multiples controverses dont elle a fait l'objet, et qui se poursuivent aujourd'hui. Le débat franc et ouvert est souvent esquivé ou refusé. Les réflexions critiques restent en partie inaudibles, car la plupart du temps cantonnées sur des supports de publication marginalisés. Certains résultats, présentés comme des « découvertes » inédites, ne sont en fait que la reprise de conclusions obtenues par des études antérieures, mais systématiquement occultées. D'autres résultats distordent des réalités complexes et aboutissent à des préconisations dangereuses.

Réinjecter de la pluralité méthodologique

Les RCT, nous l'avons dit, représentent une avancée. Mais celle-ci est limitée et s'accompagne d'une double régression : épistémologique d'abord, les promoteurs des RCT partageant une conception positiviste de la science, aujourd'hui surannée ; politiquement ensuite, par le caractère impérialiste d'une démarche prétendant comprendre tous les mécanismes de développement par cet instrument.

Déjà en 2012, R. Picciotto se demandait quand la bulle (des RCT) allait éclater. Les attaques vont croissant, et la nobélisation d'A. Deaton, fervent critique des RCT, va probablement accélérer le processus. Si les RCT restent probablement adaptées et légitimes pour certaines politiques circonscrites de manière précise, il est à la fois nécessaire et possible d'employer d'autres méthodes.

Celles-ci adoptent une approche pragmatique, définissant les questions de recherche et les outils méthodologiques nécessaires au cas par cas, en concertation avec les partenaires impliqués (opérateurs de terrain, bailleurs, etc.). Elles s'appuient également sur un pluralisme méthodologique, fondé sur l'interdisciplinarité et reconnaissant la diversité des modes d'administration de la preuve (inférence statistique/analyse compréhensive).

Il ne s'agit pas de récuser le formalisme ou la modélisation, mais d'en faire un usage contrôlé. Enfin ces approches ne visent pas l'énoncé de lois universelles, mais cherchent à expliciter des liens de causalité propres à une période et à un contexte précis.