

Prix Nobel d'économie 2019 : les limites de la méthode des essais cliniques

24 octobre 2019, 18:51 CEST • Mis à jour le 25 octobre 2019, 00:20 CEST



Annnonce de l'attribution du prix Nobel d'économie à Abhijit Banerjee, Esther Duflo et Michael Kremer (de gauche à droite sur l'écran) pendant une conférence de presse tenue à l'Académie royale des sciences de Suède à Stockholm, le 14 octobre 2019. Jonathan Nackstrand/AFP

Auteurs



Isabelle Guérin

Directrice de recherche à l'IRD-Cessma, Membre de l'Institute of Advanced Study, Princeton (2019-2020), Institut de recherche pour le développement (IRD)



François Roubaud

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



Langues

- English
- Français

Le prix Nobel d'économie vient d'être attribué à Esther Duflo, Abijit Banerjee et Michael Kremer pour leur travail consistant à adapter la méthode des essais cliniques aux interventions en matière de développement. Le jury Nobel a jugé que ce nouveau type d'expérimentation a « considérablement amélioré notre capacité à lutter contre la pauvreté globale » et « à transformer l'économie du développement ». S'il y a des raisons de s'en réjouir (l'une des trois nommé-e-s est une femme, jeune et française de surcroît ; le prix rend ses lettres de noblesse à l'économie du développement et à des travaux empiriques proches du terrain), il faut néanmoins questionner la validité et les conséquences

de l'usage croissant de cette méthode.

Réactualisant un article déjà publié, nous réitérons nos réserves. 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.

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. Cette méthode, couramment appliquée depuis le milieu du XX^e siècle dans le domaine de la médecine – où elle suscite de nombreux débats –, a ensuite été transposée à l'évaluation des politiques publiques dans les domaines de l'éducation, de la criminalité, de la fiscalité, etc., notamment aux États-Unis dans les années 1960-1980.

Depuis une quinzaine d'années, ces essais cliniques (qu'on désigne couramment par leur acronyme anglais RCT, pour *randomized control trials*) se sont ouverts à un champ nouveau : celui des politiques et de l'aide au 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 ».

À lire aussi : *Le principe des essais cliniques peut-il s'appliquer aux politiques de développement ?*

Un prétendu monopole de la rigueur scientifique

Les RCT sont présentées par leurs adeptes comme une véritable révolution copernicienne, mise en avant dans l'ouvrage *Repenser la pauvreté* d'Esther Duflo et Abijit Banerjee, ou dans leurs interventions publiques (voir également celle-ci). La communauté académique et politique tend à leur attribuer en exclusive les qualificatifs de « rigoureuses », voire de « scientifiques ».

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 par exemple à 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.

Cet effet d'éviction est-il réellement 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 rurale du Maroc s'appliquent à une autre région marocaine, à la Tunisie voisine ou encore à la Bolivie. Cet argument est classique et bien admis par tous. Ceux qui suivent le sont moins.
- Contrairement à ce qui est souvent asséné, la validité interne des RCT pose également 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'enquête 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 de parties prenantes (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. Ces bricolages se font, 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), leur notoriété, parfois acquise grâce aux promoteurs des RCT

(exemple de la controverse sur les vermifuges) et, parfois, les exigences de publication des chercheurs...



Fabrication d'une bière locale dans le cadre d'un programme de microcrédit à Leo, Burkina Faso, 24 janvier 2014. Lionel Bonaventure/AFP

Une illustration

Nous avons récemment répliqué une RCT menée par Esther Duflo et ses collègues sur le microcrédit au Maroc. Ce type d'exercice, essentiel pour garantir la fiabilité d'une étude, consiste à reprendre les données brutes de l'enquête et à en reproduire les résultats. Nous sommes parvenus à reproduire les résultats, ce qui est une bonne nouvelle, mais en mettant en évidence une multiplicité de problèmes et d'erreurs qui affectent profondément la validité interne et externe de cette RCT. En voici quelques exemples :

- Un échantillonnage très différent du protocole initial, si bien qu'on ne peut pas caractériser la population étudiée et interpréter de quoi les résultats sont représentatifs ;
- Le sexe et l'âge des membres des ménages censés avoir été interrogés avant et après l'intervention varie tellement qu'il ne peut s'agir des mêmes ménages dans 20 % des cas ;
- Des estimations incohérentes du patrimoine possédé par les ménages, alors qu'il s'agit d'une variable centrale pour évaluer l'impact économique du programme ;
- Alors que la zone d'enquête était supposée vierge de crédit avant le traitement et que la zone de

contrôle était supposée le rester pendant l'étude, ce n'est pas le cas ;

- Les chercheurs ont arbitrairement choisi d'écarter avant analyse les 27 ménages (0,5 % du total) présentant les valeurs les plus élevées sur certaines variables. Si on en écarte 12 de plus ou 12 de moins (0,3 % ou 0,7 % du total), les résultats changent totalement.

Cette réplique a donné lieu à des échanges avec Esther Duflo et ses collègues, disponibles ici sous la forme de documents de travail. Ces échanges attestent de nos profondes différences de vues sur ce qui fonde la validité scientifique d'une étude de terrain. Il nous semble nécessaire que nos pairs se penchent plus attentivement sur cette question.

Les raisons du succès

En définitive, le type d'interventions susceptibles d'être évaluées par les RCT est restreint : 5 % 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 ou encore les échanges internationaux, pour n'en citer que quelques-unes.

Comment justifier un tel succès ? Ce n'est pas toujours la supériorité scientifique de certaines méthodes ou théories qui explique leur réussite, mais la capacité de leurs promoteurs à convaincre à un moment donné un nombre suffisant d'acteurs. En d'autres termes, le succès vient à la fois d'une offre et d'une demande. Du côté de la demande, le succès des RCT illustre l'évolution de la discipline économique (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).

Le succès des RCT illustre également les transformations du secteur de l'aide au développement, où se multiplient les petits projets s'efforçant de corriger les comportements individuels plutôt que de mettre en place ou de maintenir des infrastructures et des politiques nationales de développement.

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. Ces chercheurs 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, mobilisation compassionnelle et engagement moral) 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.

Esther Duflo, Nobel française d'économie : "C'est le travail de tou...



Interview d'Esther Duflo le 15 octobre 2019, au lendemain de l'attribution du prix Nobel d'économie (Europe 1).

Les RCT appliquées au développement pourraient être une avancée scientifique, à condition d'en reconnaître les limites (nombreuses) et le champ d'application (étroit). Prétendre résoudre la pauvreté avec ce type de méthode, comme le revendiquent certain-es de ses promoteurs, et au premier chef les trois lauréats du prix Nobel, est une double régression : épistémologique d'abord, puisque cette prétention illustre une conception positiviste de la science, aujourd'hui surannée ; politique ensuite, puisque des questions pourtant centrales pour la compréhension et la lutte contre la pauvreté et les inégalités sont laissées de côté.

La consécration va-t-elle conduire les randomisateurs du développement à plus de mesure quant aux bienfaits des différentes méthodes, ou au contraire à en profiter pour consolider leur position déjà quasi hégémonique ? Il y a de bonnes raisons d'être inquiets.

Florent Bédécarrats a contribué à cet article.



Prix Nobel **pauvreté** **recherche** **comportement** **économie comportementale**