RÉSEAU-CANOPÉ.FR
CANOPÉ
LE RÉSEAU DE CRÉATION
ET D'ACCOMPAGNEMENT PÉDAGOGIQUES

Les évaluations par assignation aléatoire

Apports et limites

Arthur Jatteau

Dans Idées économiques et sociales 2018/3 (N° 193), pages 34 à 41
Éditions Réseau Canopé

ISSN 2257-5111
DOI 10.3917/idee.193.0034

Article disponible en ligne à l'adresse
https://www.cairn.info/revue-idees-economiques-et-sociales-2018-3-page-34.htm

CAIRN.INFO
MATIÈRES À RÉFLEXION

Découvrir le sommaire de ce numéro, suivre la revue par email, s'abonner...
Flashez ce QR Code pour accéder à la page de ce numéro sur Cairn.info.

Distribution électronique Cairn.info pour Réseau Canopé.
La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

© Réseau Canopé | Téléchargé le 22/08/2024 sur www.cairn.info par Théophile Courtier (IP: 91.171.117.182)

Les évaluations par assignation aléatoire : apports et limites

Les évaluations par assignation aléatoire, portées notamment par l'économiste Esther Duflo, constituent une méthode quantitative d'évaluation d'impact qui a connu un large succès ces dernières années. Quels en sont les fondements théoriques, les domaines d'application, les limites épistémologiques et méthodologiques ?

Arthur Jatteau, agrégé de sciences économiques et sociales, maître de conférences en économie et en sociologie à l'université de Lille (Clersé)

En janvier 2009, quand Esther Duflo donne la leçon inaugurale de son cours au Collège de France, qui lui a attribué une chaire annuelle, elle est âgée de seulement 36 ans. L'année suivante, elle se verra décerner la médaille John Bates Clark, largement vue comme la récompense la plus prestigieuse en économie, après le « Nobel », dont on considère qu'elle est un prélude. Ces distinctions ne constituent qu'une fraction des honneurs obtenus par l'économiste franco-américaine, professeure au MIT, à Boston, depuis une quinzaine d'années. La principale raison à cette fulgurante carrière tient à la méthode qu'elle a contribué à promouvoir à travers le monde : l'évaluation par assignation aléatoire[1]. Si elle n'en est pas l'inventrice, elle a néanmoins cofondé un laboratoire de recherche, le J-PAL (Poverty Action Lab), qui a lancé près de mille expérimentations de ce type depuis 2003, sur tous les continents. Aujourd'hui, cette méthode est reconnue par de nombreux économistes et institutions comme étant le « *gold standard* » de l'évaluation, tant dans le champ de l'économie du développement que de l'économie des politiques publiques.

Mais de quoi s'agit-il exactement ? L'évaluation par assignation aléatoire permet d'évaluer les effets d'un programme ou d'un dispositif sur un groupe d'individus en s'appuyant sur le tirage au sort d'un groupe de bénéficiaires[2] et d'un groupe de non-bénéficiaires. Cette méthode rencontre un succès fulgurant depuis qu'Esther Duflo et son équipe l'ont remise sur le devant de la scène académique dans la seconde moitié de la décennie 2000.

L'évaluation par assignation aléatoire s'applique à de nombreux domaines : l'éducation, la santé, la gouvernance, la microfinance, l'environnement, le travail… Elle cherche à répondre à des questions relativement simples : est-ce que la distribution gratuite d'uniformes scolaires à des élèves d'une région d'un pays où ils sont obligatoires mais payants favorise leur taux de présence en classe [2] ? Est-ce qu'un changement dans le message de prévention à l'égard des maladies sexuellement transmissibles a un impact sur leur diffusion [3] ? Est-ce que les audits gouvernementaux réduisent la corruption [4] ?

Quoique séduisante de prime abord, notamment grâce à son apparente simplicité, la méthode de l'évaluation par assignation aléatoire souffre de sérieuses limites dont la prise en compte s'avère nécessaire. L'engouement en sa faveur a pu laisser entendre qu'enfin, en matière évaluative, le « Graal » aurait été trouvé. Esther Duflo et Michaël Kremer affirment même que cette méthode a le potentiel pour révolutionner les politiques publiques au XXIe siècle, comme les essais cliniques randomisés l'ont fait en médecine au XXe siècle [5]. La révolution des évaluations par assignation aléatoire a-t-elle eu lieu ?

Dans un premier temps, nous présenterons la méthode elle-même, en montrant comment, par

1 *Le terme rencontre en français plusieurs autres acceptions [1] : expérimentation aléatoire, évaluation contrôlée randomisée… En anglais, une expression fait davantage consensus :* randomized controlled trial, *souvent abrégée en RCT.*

2 *Il peut même y avoir plusieurs groupes de bénéficiaires, par exemple si le programme testé prend plusieurs formes.*

© Réseau Canopé | Téléchargé le 22/08/2024 sur www.cairn.info par Théophile Courtier (IP: 91.171.117.182)

le biais du tirage au sort, elle cherche à résoudre la question fondamentale en évaluation du contrefactuel. Ensuite, nous présenterons quelques résultats issus de domaines variés. Enfin, nous présentons deux critiques qui nous semblent les plus importantes.

Une méthode astucieuse fondée sur le tirage au sort

À la rentrée scolaire de 2017, une mesure phare du ministre de l'Éducation nationale, Jean-Michel Blanquer, est entrée en application : la limitation à 12 du nombre d'élèves par classe de CP dans les réseaux d'éducation prioritaire les plus difficiles (REP +). Quoi que l'on pense d'une telle mesure, il peut être légitime d'en vouloir mesurer l'efficacité. Est-ce que la réduction du nombre d'élèves par classe dans les écoles les plus défavorisées est une politique qui marche ? Avant même de répondre à une telle question, il reste bien évidemment à s'entendre sur les critères qui permettent de juger de l'efficacité d'une telle politique [6]. C'est un débat important, mais que nous laissons de côté pour le moment pour nous centrer sur l'aspect le plus technique du problème. Prenons par exemple comme indicateur de réussite le niveau de lecture des élèves, mesuré à l'aide d'un test standardisé.

Comment, dès lors, attester des effets de la réduction du nombre d'élèves par classe ? Une première idée, fort simple, consisterait à faire passer ces tests de lecture à tous les élèves de CP en France, qu'ils soient ou non scolarisés en REP +. Il suffirait ensuite de faire la différence entre la note moyenne obtenue par les élèves scolarisés dans des classes de 12 élèves, et les autres (dont il est permis de penser que tous ou presque seront scolarisés dans des classes avec plus de 12 élèves). Si celle des premiers est supérieure à celle des seconds, alors on pourrait juger que cette politique s'est avérée efficace.

Pourtant, on saisit aisément qu'un problème important subsiste et qu'il ruine notre protocole d'évaluation : les élèves en REP +, qui sont scolarisés dans des classes de plus faibles effectifs, n'ont sans doute pas le même niveau de départ que ceux scolarisés dans des classes à effectifs plus fournis et qui ne sont pas situés en REP +. De nombreuses études statistiques et enquêtes montrent que le niveau scolaire est corrélé avec l'origine sociale. Et ce sans compter que les écoles classées REP + sont susceptibles de bénéficier de moyens supplémentaires. Comparer les élèves de CP en REP + avec les élèves de CP des écoles qui ne sont pas en REP + reviendrait à comparer des groupes qui ne sont pas comparables. Une telle opération ne serait pas instructive pour mesurer l'effet réel d'une telle réforme, car une différence entre les moyennes obtenues aux tests pourrait provenir d'un autre facteur, différent de celui du nombre d'élèves par classe. Or c'est précisément ce dernier effet que l'on veut isoler.

C'est sur ce point précis que l'évaluation par assignation aléatoire apporte un élément méthodologique crucial : le tirage au sort. Plusieurs méthodes existent pour essayer de former des groupes qui soient comparables [7], mais la plus satisfaisante sur le plan statistique est l'utilisation de l'aléatoire. Dans notre exemple il s'agirait, préalablement à la mise en place d'une telle réforme, de tirer au sort, parmi les élèves de CP en REP +, certains qui bénéficieraient de la mesure et d'autres qui n'en bénéficieraient pas. Les premiers feraient ainsi partie de ce que l'on appelle le « groupe test » (ou encore le « groupe de traitement ») et les seconds du « groupe témoin » (ou encore « groupe de contrôle »). Le tirage au sort maximise les chances que les groupes soient comparables, et cela d'autant plus que leur effectif est élevé. S'il est fort probable qu'en tirant au sort, parmi les élèves de CP en REP +, deux élèves pour le groupe test et deux élèves pour le groupe témoin, ces groupes soient très différents, ce n'est plus vrai si les groupes sont constitués de mille élèves chacun, par exemple. Dans ce cas, il y a de grandes chances (et elles se mesurent [8]) pour qu'il y ait le même taux de fils d'ouvriers, le même âge moyen, la même taille moyenne, le même niveau, etc. [3].

Ainsi, voici les étapes pour procéder à la mesure des effets de la politique limitant à 12 le nombre d'élèves dans les classes de CP en REP +. Tout d'abord, il faut constituer, par tirage au sort, un groupe test et un groupe témoin [4]. Les élèves du premier iront ainsi dans des classes limitées à 12 élèves, alors que les autres ne bénéficieront pas d'un tel dispositif. Le tirage au sort effectué en amont assure leur comparabilité. À la fin de l'année, on fait passer un test aux élèves qui font partie de l'évaluation. Il suffit alors de comparer la note moyenne obtenue par les élèves du groupe test et celle obtenue par ceux du groupe témoin [5]. La différence entre les deux permettra de mesurer les effets d'une telle politique.

3 Il est évidemment possible que subsistent des différences mais cette méthode permet de maximiser les chances pour qu'elles ne soient pas significatives.

4 Ce protocole que nous proposons se veut simplificateur. En réalité, il peut être plus pertinent de tirer au sort au niveau des classes ou au niveau des écoles [9], pour des questions tant pratiques (difficultés auprès des parents et du corps académique) que techniques (risque d'externalités positives des enfants des classes tests vers ceux des classes témoins).

5 Il est possible, pour plus de sécurité, de faire passer un tel test au début de l'année scolaire aux deux groupes, afin de s'assurer qu'ils sont bien comparables au moins en ce qui concerne cette variable, cruciale ici, qu'est le niveau scolaire. Cependant, le tirage au sort permet en toute rigueur de s'exonérer d'une telle étape, qui permet également de voir si les deux groupes ont progressé au cours de l'année.

© Réseau Canopé | Téléchargé le 22/08/2024 sur www.cairn.info par Théophile Courtier (IP: 91.171.117.182)

On le voit : la méthode de l'évaluation par assignation aléatoire est simple. Elle ne suppose pas de connaissances ardues en mathématiques ou en statistiques, comme c'est le cas pour d'autres méthodes quantitatives d'évaluation [10]. Néanmoins, elle suppose que soient constitués *ex ante* les groupes témoin et test, le tirage au sort ne pouvant intervenir *a posteriori*. C'est pourquoi, à l'exception de très rares cas [11, 12, 13], cette méthode est essentiellement utilisée dans un cadre expérimental et non pour évaluer une politique déjà lancée à grande échelle.

Trois résultats d'évaluations par assignation aléatoire

Environ un millier d'évaluations par assignation aléatoire ont été menées depuis une quinzaine d'années, sans compter celles dans les années 1970 et 1980 aux États-Unis [14]. Afin d'en présenter quelques exemples, il nous a fallu faire un choix et nous en présentons trois : dans trois domaines différents (éducation, santé, emploi), dans un pays pauvre (le Kenya) pour les deux premières et dans un pays riche (la France) pour la troisième.

La distribution gratuite de manuels scolaires permet-elle d'améliorer le niveau des élèves ?

La question des moyens revient régulièrement dans les débats concernant l'école. Peut-on faire mieux sans hausse de moyens ou en faut-il nécessairement davantage pour améliorer les choses ? Ce débat rencontre des arguments passionnés, mais dont le soubassement scientifique n'est pas toujours clair. C'est pourquoi Paul Glewwe, Michaël Kremer et Sylvie Moulin ont mis en place, au Kenya, une évaluation par assignation aléatoire afin de chercher si la distribution gratuite de manuels scolaires permettrait d'améliorer le niveau scolaire des écoliers [15]. Ils étaient en effet très peu présents à l'époque dans les écoles kényanes.

En 1995, une ONG hollandaise a financé un programme de distribution de manuels scolaires auprès de plusieurs dizaines d'écoles dans le pays. Elles furent divisées aléatoirement[6] entre plusieurs groupes. Outre un groupe témoin, plusieurs groupes tests reçurent des manuels scolaires mais à des dates différentes. Un test de niveau fut passé au début de l'année scolaire, avant la moindre distribution, et un autre à la fin de l'année scolaire, après la distribution. Le premier test permit de s'assurer qu'il n'y avait pas de différence significative entre les groupes avant que les élèves des groupes tests obtiennent les manuels scolaires, ce que la randomisation a grandement favorisé.

L'impact général peut être jugé décevant, puisque les différents groupes n'obtiennent pas de résultats significativement différents aux tests de fin d'année, ce qui veut dire que les élèves qui ont reçu un manuel scolaire n'ont pas atteint un niveau supérieur à ceux qui n'en ont pas reçu. Comment cela peut-il s'expliquer ? Les auteurs soulignent que la langue dans laquelle ils étaient rédigés, l'anglais, n'est que la troisième langue des élèves à qui ils étaient destinés. Ils parlent d'abord des langues régionales, ensuite le swahili, l'anglais n'intervenant que comme troisième langue, généralement mal maîtrisée. Il convient ainsi de prendre garde à toute conclusion hâtive que l'on pourrait tirer d'une telle évaluation concernant l'impact des manuels scolaires à l'école, et plus généralement sur les effets d'une hausse des moyens éducatifs.

Faut-il donner ou faire payer les moustiquaires ?

En économie de la santé, une question cruciale occupe les économistes : est-ce qu'il vaut mieux rendre gratuits les biens de santé (médicaments, consultation…) ou bien les faire payer, à un prix même modique ? La première option assure une diffusion et un accès le plus large possible, puisqu'il n'y a pas de barrière financière. Mais le risque, c'est la surconsommation médicale et le manque de considération pour les biens en question. C'est pourquoi certains militent pour qu'ils soient vendus, même si les prix n'ont pas vocation à couvrir l'intégralité des coûts, de manière d'une part à limiter l'accès à ceux qui en ont vraiment besoin, car ce sont eux qui seront les plus enclins à payer, et d'autre part à inciter les consommateurs à accorder de l'importance aux biens en question (médicaments, consultations chez le médecin…). Que l'on considère l'une ou l'autre option, les arguments semblent se tenir sur un strict plan théorique.

C'est pour trancher cette question d'un point de vue empirique que deux économistes, Jessica Cohen et Pascaline Dupas, ont mis en place une évaluation par assignation aléatoire [16]. Dans une région à l'ouest du Kenya, elles ont sélectionné plusieurs maternités où sont généralement distribuées des moustiquaires à des femmes enceintes

6 *En réalité, il n'y a eu pas de tirage au sort à proprement parler. Les écoles furent d'abord classées par ordre alphabétique, puis, en partant de la première, elles furent assignées toutes les quatre écoles au premier groupe, puis le même procédé fut suivi pour le second groupe, etc.*

© Réseau Canopé | Téléchargé le 22/08/2024 sur www.cairn.info par Théophile Courtier (IP: 91.171.117.182)

afin de prévenir les piqûres de moustiques, qui transmettent le paludisme. Reprenant les termes du débat que nous venons de mentionner, elles ont cherché à savoir s'il était préférable de distribuer gratuitement à toutes une moustiquaire, ou bien de les faire payer à un prix tout de même subventionné. Elles ont donc formé plusieurs groupes de cliniques, l'un servant de groupe témoin, où tout se déroulait comme habituellement, un groupe test dans lequel les moustiquaires étaient données gratuitement et plusieurs autres groupes tests où elles étaient vendues à différents prix.

Ce dispositif permet déjà de savoir si le prix constitue une barrière dans l'accès à ce bien. Mais les deux chercheuses l'ont complété afin de tester deux autres éléments intéressants. Tout d'abord, elles ont fait pratiquer aux femmes des maternités un prélèvement afin de mesurer leur taux d'hémoglobine dans le sang, qui constitue un indicateur d'exposition au paludisme. Cela permet de vérifier si le fait de mettre un prix aux moustiquaires engendre le phénomène selon lequel celles qui les acquièrent davantage sont celles qui en ont le plus besoin. Ensuite, des enquêteurs de terrain sont allés rendre visite à un échantillon des femmes chez elles, quelques mois après leur sortie de la maternité, afin de constater si elles utilisaient la moustiquaire ou non. Cette visite permettait ainsi de savoir si le prix du bien de santé avait joué sur son utilisation ultérieure.

Les résultats de Cohen et Dupas montrent que le simple fait de mettre un prix réduit drastiquement la proportion de bénéficiaires. Le passage d'un prix nul à un prix de 0,60 $ la fait baisser de 60 %, ce prix étant même inférieur à celui que les ONG de la région proposaient au moment de l'évaluation. En ce qui concerne l'usage qui est fait de ces moustiquaires, les chercheuses ne trouvent pas de différences significatives entre les différents groupes tests : un prix plus élevé ne semble pas impliquer un usage plus important. Par ailleurs, elles montrent que, dans le cadre de leur expérimentation, vendre les moustiquaires à un prix plus élevé n'implique pas que les femmes les plus exposées au paludisme les achètent davantage que celles qui y sont moins exposées. Autrement dit, faire payer ne permet pas de discriminer correctement entre les consommateurs du point de vue de leur nécessité sanitaire.

Les conclusions qu'apporte ce travail sont intéressantes, dans la mesure où elles essaient de répondre clairement, avec une méthode d'évaluation simple et en apparence efficace, à des questions cruciales en économie de la santé. Avant d'explorer plus en profondeur les différentes limites de l'évaluation par assignation aléatoire dans notre dernière partie, nous pouvons néanmoins dès maintenant souligner que les particularités de cette expérimentation posent directement la question de la généralisation possible des résultats. Elle est géographiquement circonscrite à une région du Kenya et socialement limitée à une catégorie particulière de la population : les femmes enceintes. Dès lors on peut se demander quels enseignements généraux sur la tarification des biens de santé l'on peut tirer d'une telle expérimentation, au-delà du contexte dans lequel elle s'est déroulée.

Le service public de placement des chômeurs fait-il mieux que les opérateurs privés ?

Parce que le chômage est une question lancinante dans bon nombre de pays industrialisés depuis les années 1980, en particulier en France, c'est aussi une question particulièrement vive de la théorie économique. Pourquoi existe-t-il et, surtout, comment y remédier ? Parmi les solutions dont il est question dans la littérature économique, on trouve celle du recours aux opérateurs privés de placement des chômeurs. Pôle emploi, dont c'est aujourd'hui la charge, est un service public et doit, à ce titre, répondre à des objectifs qui lui sont propres. Ce faisant, certains économistes affirment que cet organisme comme les conseillers qui y sont employés manquent d'incitations pour faire correctement leur travail. Des entreprises privées, qui seraient payées partiellement en fonction de leur réussite à (re)trouver un emploi aux chômeurs, pourraient de ce point de vue obtenir de meilleurs résultats. Le privé ferait-il mieux que le public en matière de « placement » des chômeurs ?

Pour apporter des éléments de réponses à cette question, les économistes Luc Behaghel, Bruno Crépon et Marc Gurgand ont mis en place une évaluation par assignation aléatoire [17, 18]. Parmi des chômeurs, trois groupes ont ainsi été constitués par tirage au sort. Le premier, le groupe de contrôle, est suivi par la voie classique à l'ANPE, devenue depuis Pôle emploi. Le second est assigné à un dispositif innovant mené au sein de l'ANPE, « Cap vers l'entreprise ». L'idée principale y était de réduire

© Réseau Canopé | Téléchargé le 22/08/2024 sur www.cairn.info par Théophile Courtier (IP: 91.171.117.182)

le nombre de demandeurs d'emplois suivis par les conseillers (40, contre 120 environ habituellement). Le troisième est envoyé vers des opérateurs privés de placement, avec un ratio nombre de chômeurs/conseiller amélioré dans des proportions similaires. Mais la structure de financement de ces opérateurs était originale : ils touchaient une somme forfaitaire pour chaque chômeur intégré dans le programme, et éventuellement un bonus en fonction du taux de retour à l'emploi et de la durée d'occupation de l'emploi.

Le premier résultat de cette évaluation est que la réduction du nombre de demandeurs d'emploi à gérer par conseiller est bénéfique. Les chômeurs assignés aux groupes tests ont ainsi plus de chances de retrouver un emploi que ceux assignés au groupe de contrôle. Le deuxième enseignement, qui va à l'encontre de bon nombre de résultats et d'intuitions tirés de la théorie économique, est que le public semble faire mieux que le privé en la matière. Les chômeurs du programme « Cap vers l'entreprise » ont ainsi plus de chances de retrouver un emploi, et ce plus rapidement, que ceux suivis par un opérateur privé de placement.

Bien sûr, réduire le nombre de dossiers que doivent suivre les conseillers représente un coût non négligeable, ce qui alourdit les dépenses publiques. Mais si les chômeurs sont plus nombreux à retrouver un travail et de manière plus rapide, alors cela engendre des économies en termes de prestations sociales et de coût des services de placement des chômeurs. Les auteurs de cette étude ont procédé à des calculs coûts-avantages qui tendent à montrer que le programme innovant du service public est efficient, c'est-à-dire qu'il rapporte plus qu'il ne coûte, alors que les opérateurs privés ne le sont pas.

Cette évaluation permet de pointer une nouvelle limite. Si l'on généralise de tels dispositifs à tous les chômeurs, par exemple en embauchant de nombreux conseillers de manière à faire drastiquement baisser le nombre de dossiers qu'ils ont chacun à suivre, est-ce qu'on va retrouver de tels résultats ? Cette question est connue dans la littérature sous le nom d'« effet d'équilibre ». Pour peu que le stock d'emplois soit relativement fixe, il est permis de penser que ce genre de programmes ne contribue en réalité qu'à modifier le sens de la « file d'attente » des demandeurs d'emploi : ceux qui en bénéficient passent devant les autres, mais le nombre de chômeurs dans la file d'attente n'est que peu modifié par rapport ce qu'elle aurait été en l'absence de programme. Une telle politique de l'emploi serait ainsi en réalité peu efficace, ce que semble confirmer une étude sur le sujet [19].

Deux limites importantes

On dénombre de nombreuses limites à la méthode [1]. Aussi nous sommes-nous arrêté sur les deux qui nous semblaient les plus saillantes.

La généralisation des résultats

La limite la plus fragrante de l'évaluation par assignation aléatoire ne lui est pas spécifique, et traverse toutes les sciences sociales : c'est celle de la généralisation [20], connue sous le terme de « validité externe » dans la littérature. Comment savoir si des effets mesurés lors d'une période donnée, dans une région donnée, par une équipe donnée, pourront être retrouvés à un autre moment, dans un lieu différent, mise en œuvre par d'autres personnes ? Autrement dit : dans quelle mesure les résultats « tiennent-ils » dans un contexte différent ?

Cette question est centrale. Une réponse négative à ces questions amoindrirait considérablement l'intérêt d'une telle méthode. À quoi bon savoir que les manuels scolaires ne sont pas efficaces sur les quelques dizaines d'écoles testées dans une région du Kenya ? En quoi l'évaluation des moustiquaires nous informe-t-elle sur une autre population que celle des femmes enceintes ? Cette question de la validité externe renvoie à celle de la représentativité, sans la recouvrir exactement. Si le territoire et les individus inclus dans l'expérimentation sont représentatifs de celui et de ceux qui seraient les destinataires d'une telle politique, alors une telle méthode prend tout son sens. Mais pour arriver à une telle représentativité, il faudrait, si l'on suit la logique propre aux évaluations par assignation aléatoire, tirer au sort parmi toute la population, en prenant en compte l'inégale répartition sur le territoire. C'est techniquement impossible dans l'immense majorité des cas, en particulier dans les pays pauvres où les statistiques publiques sont souvent peu fiables.

De fait, en toute rigueur et en l'absence d'éléments affirmant le contraire, il faut considérer que les résultats des évaluations par assignation aléatoire ne sont valables que dans le contexte dans lequel elles ont été réalisées, et guère plus. C'est ce que souligne l'économiste Dani Rodrik quand il commente l'étude

© Réseau Canopé | Téléchargé le 22/08/2024 sur www.cairn.info par Théophile Courtier (IP: 91.171.117.182)

de Jessica Cohen et Pascaline Dupas sur les moustiquaires [21]. La zone où elles interviennent est très particulière, dans la mesure où de nombreuses ONG y ont fait, pendant des années, un travail de pédagogie autour des bienfaits des moustiquaires. Le terrain n'est donc pas totalement « vierge » de ce point de vue, et il est permis de penser que les résultats puissent être différents ailleurs, là où les travailleurs humanitaires sont moins présents. De plus, les chercheuses se sont centrées sur un groupe d'individus bien particulier, les femmes enceintes, qui sont les sujets les plus à risque et dont on peut penser qu'elles sont les plus sensibilisées. Au final, il est permis de se demander ce que nous apprend une telle évaluation au-delà de ce cadre, très particulier, dans lequel elle s'inscrit.

Certaines évaluations aléatoires arrivent en partie à se soustraire à un tel problème, en construisant leurs groupes sur un échantillon largement représentatif de la population potentiellement destinataire de la politique. L'évaluation de « Cap vers l'entreprise » et des opérateurs privés de placement en est un bon exemple, puisque les chômeurs inclus dans le dispositif venaient de 10 régions différentes, avec environ 400 agences locales de l'ANPE. La question de la généralisation se pose donc ici avec moins d'acuité – ce qui n'enlève rien au problème d'équilibre général que nous évoquions plus haut.

Mais la vaste majorité des évaluations aléatoires reposent sur un contexte trop particulier pour qu'il soit possible d'en tirer des enseignements généraux, ce que certains économistes font néanmoins par mégarde [22]. Il faut souligner qu'il n'y a pas de solution évidente à un tel problème. Esther Duflo et Abhijit Banerjee proposent de répliquer les dispositifs expérimentaux dans un autre contexte [23]. Cela permettrait de tester la robustesse des programmes testés en regardant si les résultats vont dans le même sens.

Pourtant, une telle solution apparaît peu convaincante. En effet, on peut se demander en quoi des résultats similaires nous informeraient sur un autre contexte non encore testé. De plus, l'intérêt de reproduire telle quelle une évaluation n'est pas toujours évident, car intervenir auprès de populations nécessite toujours une adaptation au contexte. Une piste plus prometteuse nous semble être l'étude fine des particularités des terrains dans lesquels s'inscrivent les évaluations aléatoires par le biais de la méthode de l'observation ethnographique. Cette dernière pourrait utilement compléter la démarche promulguée par Esther Duflo et ses collègues, qui se bornent aujourd'hui à une approche purement quantitative.

> *“Il faut considérer que les résultats des évaluations par assignation aléatoire ne sont valables que dans le contexte dans lequel elles ont été réalisées”*

Efficacité ou causalité ?

Une limite importante de l'évaluation par assignation aléatoire concerne également la nature de la preuve qu'elle permet de produire, et dont la perception n'a rien d'évident. Reprenons l'exemple que nous développions en introduction sur la baisse du nombre d'élèves par classe. Il est évidemment important de savoir si une telle politique marche ou pas. Mais c'est une autre question – tout aussi importante au demeurant – de savoir pourquoi elle marche (ou pas) [24]. Il s'agit de distinguer ici les preuves d'efficacité des preuves de causalité [25]. Celles-ci sont disjointes et doivent être pensées comme telles : on peut montrer qu'un dispositif « marche » sans pour autant connaître les mécanismes qui font qu'il marche [26]. C'est notamment le cas en médecine dans certaines pratiques thérapeutiques, comme

© Réseau Canopé | Téléchargé le 22/08/2024 sur www.cairn.info par Théophile Courtier (IP: 91.171.117.182)

l'acupuncture en complément des traitements contre le cancer.

Ainsi, admettons qu'une évaluation d'un dispositif visant à réduire le nombre d'élèves par classe montre que le groupe test obtient de meilleurs résultats de fin d'année que le groupe témoin. On peut en conclure que dans le cadre où elle s'est déroulée, une telle mesure est efficace. Mais pourquoi l'est-elle ? Est-ce que parce que l'instituteur peut individualiser les explications qu'il donne ? Est-ce parce qu'il peut mettre les élèves seuls chacun à sa table ? Est-ce parce qu'il y a moins de bavardages ? moins de bruit ? moins de fatigue ? Les explications sont multiples. Mais si les évaluations aléatoires peuvent donner des pistes, elles ne peuvent donner les mécanismes causaux à l'œuvre.

Cette méthode s'apparente à la mise en avant d'une corrélation, sans nécessairement mettre au jour une causalité. Quand on distribue gratuitement les moustiquaires, les femmes enceintes d'une région à l'ouest du Kenya sont mieux protégées contre le paludisme : c'est une corrélation. Savoir expliquer pourquoi il en est ainsi relève de la recherche d'une causalité, pour laquelle les évaluations aléatoires ne sont pas pertinentes. Bien sûr, il est possible de formuler des hypothèses, d'avoir des intuitions à la lecture des données ou des retours qui parviennent du terrain.

La question de la causalité est, au même titre que celle de la généralisation, une des grandes questions des sciences sociales. Étonnamment, elle peut être vue comme un impensé de l'évaluation par assignation aléatoire telle qu'elle est pratiquée depuis une quinzaine d'années, à cause d'une confusion entre preuves d'efficacité et preuves de causalité.

Pour mettre au jour ces dernières, une piste à suivre consisterait peut être à s'appuyer sur un cadre théorique particulier, comme l'ont d'ailleurs fait récemment certains économistes pratiquant les expérimentations aléatoires. Un tel cadre permet justement de tenter d'expliquer ce qui est montré par l'évaluation aléatoire. Une autre voie, trop peu rarement suivie, consiste à enquêter sur le terrain dans lequel elle s'inscrit, de manière plus profonde et plus élaborée que par une seule approche quantitative. Il s'agit ici de mobiliser le large éventail méthodologique des sciences sociales [27, 28], en ayant recours à l'observation comme déjà mentionné, mais également à l'entretien. Cela permettrait de donner du sens aux données recueillies. Dans l'exemple donné initialement, il pourrait ainsi sembler pertinent de réaliser des entretiens avec des instituteurs, des élèves mais également des parents d'élèves, des assistants d'éducation, etc. Des observations au sein des classes, mais aussi dans la salle des professeurs, pourraient également être mises à profit. C'est l'ensemble de ces matériaux, associés aux données quantitatives, qui permettrait précisément de donner du sens, et du sens causal, aux résultats mis au jour des évaluations par assignation aléatoire.

Conclusion

Assurément, la méthode de l'évaluation par assignation aléatoire est séduisante, car simple et robuste. Simple, elle assure que tout le monde comprenne comment on est parvenu aux résultats, là où le haut niveau de formalisation et de mathématisation de bon nombre d'évaluations de l'économie contemporaine exclut beaucoup de lecteurs potentiels [29]. Robuste, parce que le tirage au sort permet de maximiser les chances d'obtenir des groupes qui soient comparables.

Pour autant, cette méthode souffre de nombreuses limites qui, si elles n'en annulent aucunement l'intérêt, nuancent la prétention de supériorité arguée par ses promoteurs [30]. Certaines n'ont pas été développées ici : la transposition de la théorie au terrain (avec les effets Hawthorne, par exemple), l'impossibilité de saisir l'hétérogénéité des effets, la différence de temporalités entre les chercheurs et les politiques qui complique la prise en compte des résultats [1]… Les questions de la généralisation et de la causalité nous semblent parmi les plus importantes, d'autant qu'elles ne rencontrent pas pour le moment de réponses totalement satisfaisantes.

À cet égard, la révolution des évaluations aléatoires n'a pas (encore ?) eu lieu. Les vertus du tirage au sort, réelles, ne sauraient faire oublier leurs limites notables. Elles n'ont pas, jusqu'à maintenant, contribué à refonder les politiques publiques, que ce soit dans les pays riches ou dans les pays pauvres [7]. Cela ne doit pas contribuer à écarter pour autant l'utilisation du tirage au sort à des fins évaluatives. Mais il convient de replacer l'évaluation par assignation aléatoire dans le champ plus large des sciences sociales, afin d'en saisir tous les enjeux, en particulier concernant la généralisation et la causalité, qui constituent ses deux principaux points faibles.

7 En partie pour des raisons qui ne sont pas propres à la méthode elle-même, mais qui renvoient aux ressorts même de l'action publique [31, 32]. Aussi, croire que le caractère hautement scientifique de la méthode de l'évaluation par assignation aléatoire suffirait à sa prise en compte par les politiques relève d'une certaine naïveté.

© Réseau Canopé | Téléchargé le 22/08/2024 sur www.cairn.info par Théophile Courtier (IP: 91.171.117.182)

Bibliographie

[1] **Jatteau A.**, *Les Expérimentations aléatoires en économie*, Paris, La Découverte, 2013.
[2] **Evans D., Kremer M., Ngatia M.**, « The impact of distributing school uniforms on children's education in Kenya », *Working Paper*, 2009.
[3] **Dupas P.**, « Do teenagers respond to HIV risk information? Evidence from a field experiment in Kenya », *American Economic Journal: Applied Economics*, vol. 3, nº 1, 2011, p. 1-34.
[4] **Olken B. A.**, « Monitoring corruption: Evidence from a field experiment in Indonesia », *Journal of Political Economy*, vol. 115, nº 2, 2007, p. 200-249.
[5] **Duflo E., Kremer M.**, « Use of randomization in the evaluation of development effectiveness », in **W. R. Easterly** (dir.), *Reinventing Foreign Aid*, Cambridge, MIT Press, 2008, p. 93-120.
[6] **Desrosières A.**, *L'Argument statistique, I. Pour une sociologie historique de la quantification*, Paris, Presses des Mines, 2008.
[7] **Angrist J. D., Pischke J.-S.**, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton, Princeton University Press, 2009.
[8] **Glennerster R., Takavarasha K.**, *Running Randomized Evaluations: A Practical Guide*, Princeton, Princeton University Press, 2013.
[9] **Miguel E., Kremer M.**, « Worms: Identifying impacts on education and health in the presence of treatment externalities », *Econometrica*, vol. 72, nº 1, 2004, p. 159-217.
[10] **Fougère D.**, « Les méthodes économétriques d'évaluation », *Revue française des affaires sociales*, vol. 1, nºs 1-2, 2010, p. 105-128.
[11] **Sacerdote B.**, « Peer effects with random assignment: results for Dartmouth roommates », *The Quarterly Journal of Economics*, vol. 116, nº 2, 2001, p. 681-704.
[12] **Chattopadhyay R., Duflo E.**, « Women as policy makers: evidence from a randomized policy experiment in India », *Econometrica*, vol. 72, nº 5, 2004, p. 1409-1443.
[13] **Ferraz C., Finan F.**, « Exposing corrupt politicians: The effects of Brazil's publicly released audits on electoral outcomes », *The Quarterly Journal of Economics*, vol. 123, nº 2, 2008, p. 703-745.
[14] **Gueron J. M., Rolston H.**, *Fighting for Reliable Evidence*, New York, Russell Sage Foundation Publications, 2013.
[15] **Glewwe P., Kremer M., Moulin S.**, « Many children left behind? Textbooks and test scores in Kenya », *American Economic Journal: Applied Economics*, vol. 1, nº 1, 2009, p. 112-135.
[16] **Cohen J., Dupas P.**, « Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment », *Quarterly Journal of Economics*, vol. 125, nº 1, 2010, p. 1-45.
[17] **Behaghel L.**, « L'accompagnement personnalisé des demandeurs d'emploi », *Revue française d'économie*, vol. 28, nº 1, 2013, p. 123-158.
[18] **Behaghel L., Crépon B., Gurgand M.**, « Private and public provision of counseling to job seekers: Evidence from a large controlled experiment », *American Economic Journal: Applied Economics*, vol. 6, nº 4, 2014, p. 142-174.
[19] **Crépon B., Duflo E., Gurgand M., Rathelot R., Zamora P.**, « Do labor market policies have displacement effects? Evidence from a clustered randomized experiment », *The Quarterly Journal of Economics*, vol. 128, nº 2, 2013, p. 531-580.
[20] **Desveaux E., Fornel M.** (Éd.), *Faire des sciences sociales. Généraliser*, Paris, EHESS, 2012.
[21] **Rodrik D.**, « The new development economics: we shall experiment, but how shall we learn? », *HKS Research Working Paper*, vol. 55, 2008.
[22] **Jatteau A.**, *Les Expérimentations aléatoires en économie du développement : enjeux critiques, pratique de terrain et portée politique*, mémoire de master 2, Paris, EHESS, 2011.
[23] **Banerjee A. V., Duflo E.**, « L'approche expérimentale en économie du développement », *Revue d'économie politique*, vol. 119, 2009, p. 691-726.
[24] **Deaton A.**, « Instruments, randomization, and learning about development », *Journal of Economic Literature*, vol. 48, nº 2, 2010, p. 424-455.
[25] **Laurent C.** *et al.,* « Pourquoi s'intéresser à la notion d'"*evidence-based policy*" ? », *Tiers Monde*, nº 200, 2009, p. 853-873.
[26] **Labrousse A.**, « Nouvelle économie du développement et essais cliniques randomisés : une mise en perspective d'un outil de preuve et de gouvernement », *Revue de la régulation*, nº 7, 2010, p. 2-32.
[27] **Olivier De Sardan J.-P.**, *La Rigueur du qualitatif. Les contraintes empiriques de l'interprétation socio-anthropologique*, Louvain-la-Neuve, Academia Bruylant, 2008.
[28] **Jatteau A.**, *Faire preuve par le chiffre ? Le cas des expérimentations aléatoires en économie,* thèse, Saclay, ENS Paris Saclay, 2016.
[29] **Imbens G. W., Wooldridge J. M.**, « Recent developments in the econometrics of program evaluation », *Journal of Economic Literature*, vol. 47, nº 1, 2009, p. 5-86.
[30] **Banerjee A. V.** (dir.), *Making Aid Work*, Cambridge, The MIT Press, 2007.
[31] **Weiss C. H.**, « The circuitry of enlightenment: Diffusion of social science research to policymakers », *Science Communication*, vol. 8, nº 2, 1986, p. 274-281.
[32] **Delarue J., Naudet J.-D., Sauvat V.**, « Les évaluations sont-elles utiles ? Revue de littérature sur "connaissances et décisions" », *Ex Post*, nº 3, 2009.

© Réseau Canopé | Téléchargé le 22/08/2024 sur www.cairn.info par Théophile Courtier (IP: 91.171.117.182)