Philosophia Scientiæ

Travaux d'histoire et de philosophie des sciences

23-2 | 2019

Expérimentation dans les sciences de la nature
Expérimentation dans les sciences humaines et
sociales

Les essais contrôlés randomisés. Une comparaison entre la médecine et l'économie

Randomized Controlled Trials. A comparison between medicine and economics

Arthur Jatteau

Édition électronique

URL : <http://journals.openedition.org/philosophiascientiae/1933>

DOI : 10.4000/philosophiascientiae.1933

ISSN : 1775-4283

Éditeur

Éditions Kimé

Édition imprimée

Date de publication : 24 mai 2019

Pagination : 85-110

ISBN : 978-2-84174-933-1

ISSN : 1281-2463

Référence électronique

Arthur Jatteau, « Les essais contrôlés randomisés. Une comparaison entre la médecine et l'économie », *Philosophia Scientiæ* [En ligne], 23-2 | 2019, mis en ligne le 01 janvier 2021, consulté le 30 mars 2021. URL : <http://journals.openedition.org/philosophiascientiae/1933> ; DOI : <https://doi.org/10.4000/philosophiascientiae.1933>

Tous droits réservés

Les essais contrôlés randomisés. Une comparaison entre la médecine et l'économie

Arthur Jatteau

Université de Lille,
CLERSÉ, UMR 8019 –
ENS Paris Saclay, IDHE.S, UMR 8533 (France)

Résumé : Depuis le début des années 2000, les essais randomisés contrôlés ont connu un fulgurant retour sur le devant de la scène académique dans le domaine de l'économie du développement et plus généralement dans celui de l'évaluation des politiques publiques. Alors que ses partisans revendiquent une filiation avec les essais cliniques randomisés en médecine, ces derniers demeurent mal connus par les économistes. Cet article vise à combler ce manque en proposant une comparaison de l'usage et des limites des essais contrôlés randomisés entre la médecine et l'économie. Nous montrons ainsi que si cette méthode est bien souvent placée au sommet de la hiérarchie des preuves en médecine, des voix critiques existent et méritent d'être prises en compte pour penser son application en économie. Nous soulignons également que le traitement du problème de la généralisation des résultats, très bien documenté en médecine, permet de nuancer l'intérêt de la méthode en économie.

Abstract: Since the beginning of the 2000s, Randomized Controlled Trials (RCTs) have returned to center stage, both in the economics of development and in the evaluation of public policy. Despite the fact that RCT promoters claim a filiation with medicine, these economists seem unfamiliar with the literature around clinical trials. This article aims to overcome this limitation, by comparing the use and the limits of RCTs between medicine and economics. I show that while this method is considered to be the “gold standard” of evidence in medicine, some limits exist and have to be taken into account in economics. I also show how the way in which the problem of external validity is addressed in medicine provides arguments to obtain a more nuanced picture of the interest of RCTs in economics.

Depuis le début des années 2000, une méthode d'évaluation d'impact, basée sur le tirage au sort de groupes (témoin et test), connaît un vif succès en économie, en particulier dans le champ du développement : les expérimentations aléatoires¹ [Duflo 2010*b,a*]. Méthodologiquement similaire aux essais cliniques randomisés en médecine, elle s'appuie sur le tirage au sort pour former des groupes semblables, dont l'un jouera le rôle de « contrôle » afin d'être comparé au(x) groupe(s) test(s), à qui sera distribué le « traitement » dont on souhaite mesurer l'effet. Cette méthode avait déjà été utilisée pour évaluer les politiques publiques aux États-Unis dans les années 1960-1970, dans des proportions massives [Gueron & Rolston 2013]. Les résultats furent en partie jugés décevants par les décideurs publics, en particulier à cause de leurs difficultés d'interprétation et de généralisation [Monnier 1992]. La méthode des essais contrôlés randomisés est donc passée de mode, sans toutefois disparaître [Gueron 2002].

Elle est néanmoins réapparue sur le devant de la scène en économie depuis une quinzaine d'années, par le biais de la figure montante de l'économiste franco-américaine, Esther Duflo, professeure au MIT et co-fondatrice du laboratoire de recherche J-PAL (Poverty Action Lab). Ce dernier a lancé près d'un millier d'expérimentations aléatoires depuis sa création en 2003. Cette méthode est ainsi régulièrement présentée comme le nouveau « *gold standard* » [Banerjee 2007], tant en matière d'évaluation que de lutte contre la pauvreté. Enfin, une approche permettrait d'apporter des arguments scientifiquement robustes aux meilleures voies à suivre en la matière. Les articles de recherche relatant des expérimentations aléatoires sont aujourd'hui publiés dans les revues les plus cotées de la science économique [Duflo, Kremer *et al.* 2008], [Thornton 2008], [Ashraf, Berry *et al.* 2010].

La liste impressionnante de récompenses obtenues par Esther Duflo atteste de ce succès. Le « Nobel » d'économie semble être la seule qui ne lui est pas encore attribuée, alors que son nom revient avec insistance à chaque remise de ce prix. En 2017, il a d'ailleurs été décerné à Richard Thaler pour ses travaux en finance expérimentale, quinze ans après celui obtenu par Daniel Kahneman et Vernon Smith pour leurs recherches en économie expérimentale.

Les raisons de ce succès sont diverses : érosion du consensus de Washington, crise économique, doutes sur l'efficacité de l'aide publique au développement, contraction des dépenses publiques, retour de l'empirie dans la science économique... Mais il en est une qui a permis de contribuer à la légitimation de sa méthode : la revendication d'une filiation avec la médecine [Banerjee 2007]. Il s'agissait pour les promoteurs d'une telle méthode en économie de profiter

1. Ce terme ne rencontre pas d'acception pleinement stabilisée en français. On trouve « évaluation par assignation aléatoire », « expérimentations contrôlées randomisées », « expérimentations sociales », etc. Tous recouvrent une même méthode d'évaluation qui s'appuie sur la constitution aléatoire de groupes. Nous privilégierons le terme d'expérimentation aléatoire pour les travaux en économie, celui d'essai clinique randomisé en médecine et d'essai contrôlé randomisé quand nous ne souhaitons pas les inscrire dans une discipline en particulier.

de son aura de scientificité en insistant sur le versant « médecine des preuves » [Marks 1997]. Un tel prolongement de qui est vu comme la transposition en économie d'une méthode centrale de l'« *Evidence-Based Medicine* » (EBM) est d'ailleurs salué par une grande revue médicale britannique dans un éditorial daté de 2004 [The Lancet 2004]. En s'appropriant la randomisation, c'est-à-dire l'utilisation de l'aléatoire à des fins évaluatives, l'économie deviendrait *aussi* scientifique que la médecine, selon le schéma intellectuel défendu par les économistes qui la promeuvent.

Un tel rapprochement avec la médecine mérite d'être questionné. Du point de vue historique d'abord, une recherche originale montre que l'histoire des essais contrôlés randomisés est largement pluridisciplinaire et que la médecine ne saurait prétendre à une quelconque primauté [Jatteau 2016]. La psychologie, les sciences de l'éducation, l'agronomie, la science politique, en sont au moins autant à l'origine. Du point de vue proprement épistémologique, cette filiation peut être plus classiquement contestée par ceux qui pensent l'économie comme une science sociale. À ce titre, elle ne relèverait pas du même registre qu'une science expérimentale comme la médecine [Coriat, Coutrot *et al.* 2017]. Pour les économistes des expérimentations aléatoires², il s'agit en effet d'ancrer l'économie du côté des sciences « dures », adjectif que certains d'entre eux emploient d'ailleurs sans guillemets [L'Horty & Petit 2011], par opposition à des sciences qui seraient « molles », comme les sciences sociales que la pratique actuelle des expérimentations aléatoires – et au-delà une bonne partie de la science économique contemporaine –, renie. C'est ce que Frédéric Lordon appelle le « désir de "faire science" » [Lordon 1997]. Il s'agit donc de conserver une certaine distance avec l'argumentation développée par les randomistes pour défendre et promouvoir les essais contrôlés randomisés en médecine. Certes, la filiation avec la médecine n'a que peu de fondement du point de vue de l'histoire des sciences, mais elle demeure intéressante à relever dans une optique sociologique de justification.

La revendication d'un tel héritage, par les effets performatifs qu'il entraîne nécessairement, notamment un effet de fascination très net sur les décideurs publics et sur les financeurs, mérite d'être prise au sérieux. En effet, ce parallèle artificiel que font les randomistes avec la médecine explique une partie de leur succès, outre les raisons déjà évoquées [Picciotto 2012]. Ensuite, parce qu'une étude serrée de l'utilisation des essais cliniques randomisés en médecine est à même d'éclairer leur transposition en économie. C'est l'hypothèse de travail de cet article. En effet, comme l'a noté précocement Agnès Labrousse, ce transfert de méthode cherche à agir comme un transfert de scientificité [Labrousse 2010]. L'objet des lignes qui vont suivre est de montrer que le chemin n'a été fait qu'à moitié, dans la mesure où ce transfert s'est opéré sans transfert des débats existant en médecine sur les essais cliniques randomisés. C'est un transfert qui

2. Pour souligner leur attrait pour la randomisation, nous appellerons ces économistes, s'inspirant de Deaton [Deaton 2010], les « randomistes ».

se veut partiel et partial, en particulier en ce qui concerne les limites propres à la méthode en médecine, qui sont largement laissées de côté.

Il s'agit donc pour nous ici de revenir plus précisément sur les conditions de ce transfert, en menant une recherche comparative entre la médecine et l'économie du point de vue de la méthode des essais contrôlés randomisés. Les essais cliniques randomisés étant utilisés en médecine depuis la fin des années 1940 [Armitage 2003], la littérature à ce sujet est bien plus conséquente que celle autour des expérimentations aléatoires en économie. L'expérience de leur pratique n'est pas la même non plus. La médecine a accumulé à la fois plus de connaissances, tant théoriques qu'empiriques, et plus d'expériences sur l'usage de la randomisation. Il est dès lors intéressant de s'en inspirer, en essayant d'en tirer toute la substance utile à une meilleure appréciation des expérimentations aléatoires en économie, sans pour autant croire que tous les problèmes sont identiques de part et d'autre. Précisons que notre propos se restreint volontairement aux expérimentations qui utilisent à la fois un groupe de contrôle et un tirage au sort (qui constituent le cœur des essais contrôlés randomisés), et qu'il ne saurait embrasser l'ensemble des expérimentations, en particulier celles se déroulant en laboratoire [Eber & Willinger 2012], sujet qui mériterait plusieurs autres articles, tant la littérature est abondante et ancienne [Levitt & List 2009]. Deux points justifient que l'on centre la réflexion sur les essais contrôlés randomisés *in vivo* : leur omniprésence dans le domaine de l'évaluation des projets de développement [Rodrik 2008] et leur domination académique [Jatteau 2018].

À notre connaissance, un travail comparatif entre l'usage de ces essais en médecine d'une part et en économie de l'autre n'a pas été fait jusqu'à maintenant. La médecine a peu regardé du côté de l'économie, ce qui se comprend aisément quand on sait l'antériorité de son usage massif de la randomisation. L'économie ne s'est que peu penchée sur les travaux utilisant également la randomisation en médecine. Les rares articles qui existent sur le sujet portent sur des points précis [Bruhn & McKenzie 2009], [Bulte, Pan *et al.* 2012], [Eble, Boone *et al.* 2014], mais aucun n'adopte une posture que l'on pourrait qualifier d'épistémologie comparée. Une approche exhaustive qui viserait à lister les différentes limites établies en médecine et applicables en économie dépasserait largement le cadre de cet article³. Il nous a semblé plus pertinent de se centrer sur les plus importantes. Cela implique donc d'en laisser de côté un certain nombre, en particulier celles relevant des difficultés de mise en œuvre d'une telle méthode (respect du secret de la séquence d'assignation [Schulz 1995], [Schulz & Grimes 2002], placebo, double-aveugle...) et d'autres mettant au jour le rôle des acteurs participant aux essais cliniques randomisés [Abraham & Reed 2002], [Dodier & Barbot 2000].

Les questionnements quant à la place des essais cliniques randomisés dans la hiérarchie des preuves nous semblent en effet de première importance. S'ils y figurent généralement au sommet dans le cadre de la « médecine des preuves »,

3. Sur ce point, on peut se référer au chapitre 1 de la thèse [Jatteau 2016].

des débats existent et permettent d'éclairer la supériorité revendiquée par les randomistes en économie. Cette discussion de la hiérarchie des preuves fait écho à la valorisation de méthodes autres que la randomisation afin de mesurer des effets. Un autre point central des débats est la question de la généralisation des résultats obtenus par une telle méthode, connue sous le terme de « validité externe ». Elle renvoie à la nature des savoirs ainsi obtenus, à leur application à d'autres contextes et à leur potentielle cumulativité. La validité externe demeure un problème tout aussi aigu en médecine qu'en économie. Là aussi, sa caractérisation et les pistes de solution apportées par et pour la première peuvent utilement éclairer sa transposition à la seconde.

1 La remise en cause de la supériorité des essais contrôlés randomisés

1.1 Une hiérarchie des preuves en débat

Dans le cadre de l'*Evidence Based-Medicine*, souvent traduit en français par « médecine des preuves », les essais cliniques randomisés sont généralement considérés comme le plus haut niveau de preuves [Charlton & Miles 1998], [Concato, Shah *et al.* 2000], [SIGN 50 2008]. Ce qui nous intéresse ici est d'interroger la transposition de cette hiérarchie en économie, et en particulier dans l'étude des politiques publiques, à travers le mouvement de l'*Evidence Based-Policy* (EBP), dans lequel s'inscrivent les randomistes [Laurent, Baudry *et al.* 2009].

Les randomistes reprennent à leur compte de manière explicite cette hiérarchie en soulignant la supériorité des expérimentations aléatoires sur toute autre méthode d'évaluation [Banerjee 2007]. À la supériorité des essais cliniques en médecine correspondrait celle des expérimentations aléatoires dans l'évaluation de programmes ou de projets publics. Sur ce point, soulignons que les randomistes franchissent une étape supplémentaire, ce que se gardent bien de faire la plupart des tenants de l'EBM. En effet, l'argumentation des premiers tend à délégitimer toute autre approche que la randomisation. Cette dernière serait la seule véritablement scientifique, les études quantitatives non-randomisées ou les études observationnelles ne relevant pas du domaine scientifique. Plus qu'une hiérarchie, il convient de parler chez les randomistes de monisme méthodologique [Labrousse 2016], où une seule méthode est reconnue comme étant pleinement scientifique. Fait rare dans la recherche, le laboratoire d'Esther Duflo, le J-PAL, leader dans le champ des expérimentations aléatoires, ne pratique *que* cette méthode. Tous les projets qu'ils lancent doivent suivre ainsi un même protocole s'appuyant sur le tirage au sort de groupes.

Or, il convient de préciser qu'une hiérarchie ne suppose pas une négation des méthodes qui n'occuperaient pas la première place et rend donc possible

leur utilisation en même temps qu'elle reconnaît leur utilité. C'est le cas en médecine, notamment lorsqu'il n'est pas possible, pour des raisons éthiques ou techniques, de mettre en place un essai clinique randomisé. Le recours aux autres méthodes est alors légitimé. C'est une première différence entre l'usage de la randomisation en médecine et en économie.

Si cette hiérarchie semble solidement installée en médecine, il n'en reste pas moins qu'elle suscite de nombreux débats, parfois dans des termes violents (certains allant jusqu'à parler de « fascisme » de l'*Evidence-Based Discourse* [Holmes, Murray *et al.* 2006]. Des chercheurs remettent ainsi directement en cause la randomisation et mettent en garde contre leur glorification [Feinstein 1984].

La raison principale est que les essais cliniques randomisés ne sauraient être employés pour tout de manière indifférenciée. Pour certaines pathologies par exemple, c'est souvent impossible ou très compliqué à réaliser, comme en chirurgie [Abel & Koch 1999]. Dans le cas des cancers du poumon, Michael Scriven note que l'on n'a jamais établi le lien avec le tabac à l'aide d'essais cliniques randomisés, pour des raisons éthiques, mais aussi parce que cela s'avère plus délicat pour des maladies à étiologie multifactorielle [Scriven 2008]. Pour autant, on a su prouver ce lien en l'absence de randomisation avec un degré de certitude tel qu'aucun chercheur ne s'aventurerait à remettre en cause le rôle du tabac dans ces cancers. Gordon Smith & Jill Pell traitent la question avec humour en évoquant la « dramatique » absence d'essais cliniques randomisés à même de montrer les bienfaits sur la santé de l'usage d'un parachute lors d'une chute libre [Smith & Pell 2003]. Bien en peine de trouver ne serait-ce qu'un seul essai à ce sujet, les auteurs critiquent dans leur conclusion les tenants du « tout essai clinique randomisé », et plus généralement les partisans de l'EBM, selon lesquels on ne peut rien tirer d'une étude observationnelle. En médecine, il existe donc une « zone grise » dans laquelle les canons de l'EBM, en particulier la randomisation, peuvent difficilement opérer, lorsque ce qui intéresse les chercheurs est peu testable et *a fortiori* impossible à prouver [Naylor 1995].

Or, les raisons de l'existence d'une telle « zone grise » sont globalement ignorées par les économistes qui pratiquent des essais contrôlés randomisés aujourd'hui, s'appuyant sur une vision simplifiée de l'état de la littérature autour de ces essais en médecine. Des expérimentations aléatoires sont alors menées au sujet de politiques dont les bienfaits sont unanimement reconnus depuis longtemps, comme la distribution gratuite de vermifuges contre les vers intestinaux [Miguel & Kremer 2004]. Dans certains cas, les résultats apportés par ces essais contrôlés randomisés viennent simplement confirmer des connaissances pourtant considérées comme acquises par les acteurs du développement [Quentin & Guérin 2013].

1.2 La valorisation d'autres methodes

Des chercheurs en médecine invitent ainsi à reconsidérer les méthodes autres que les essais cliniques randomisés, comme les méthodes observationnelles [Black 1996]. De telles méthodes seraient ainsi complémentaires plutôt que concurrentes. Ainsi, en économie comme en médecine, les critiques de la suprématie des essais cliniques randomisés prônent une idée simple : plutôt que de présupposer la supériorité de la méthode randomisée, faisons du cas par cas, quelle que soit la discipline considérée [Cartwright 2007]. Ils ont bien sûr leur place, mais celle-ci ne se situe pas nécessairement au-dessus des autres méthodes. D'autres chercheurs vont eux combiner des approches quantitatives (en l'occurrence des essais cliniques randomisés) et des approches qualitatives, afin d'améliorer la compréhension des effets qui touchent le groupe de traitement [Thomas, Harden *et al.* 2004].

De fait, de nombreux traitements (médicaments, techniques...) sont aujourd'hui utilisés en médecine sans que leur efficacité ait été démontrée à l'aide d'essais cliniques randomisés [Worrall 2007], ce qui n'empêche pas leur reconnaissance par la communauté médicale. Aussi, les essais cliniques randomisés ne constituent un point de passage ni nécessaire, ni suffisant, en médecine, au regard de la littérature et de la pratique médicales. La prise en compte de ces dernières doit inviter les randomistes en économie à s'interroger sur le statut de cette méthode.

Cela nous amène à considérer la prise en compte de l'EBM dans la clinique des médecins. En prescrivant certains traitements dont l'efficacité n'a pas été prouvée par des essais cliniques randomisés, beaucoup reconnaissent en effet implicitement que d'autres modes d'administration de la preuve que la randomisation sont acceptables. L'interprétation de la hiérarchie des preuves que font les praticiens n'est pas si étroite que prônent les tenants de l'EBM. Les patients qu'ils ont en face d'eux ne sauraient être réductibles aux moyennes auxquelles aboutissent ces essais. Dès lors, les praticiens doivent mobiliser d'autres modes d'administration de la preuve comme autant de guides thérapeutiques.

Dans la littérature, il existe ainsi des situations pour lesquelles un essai clinique randomisé ne s'impose pas, voire serait même contraire à l'éthique. L'étude de la nocivité de certains aliments ou produits en rend le recours délicat : il paraît inconcevable de monter un essai où le « traitement » consisterait à consommer de la drogue ou de l'alcool, pour en mesurer les effets sur l'organisme... Lorsque l'effet du traitement est très élevé se pose également la question de l'utilisation (et de l'utilité) des essais randomisés [Henry-Amar 2005]. En la matière, des débats parfois violents ont lieu concernant les médicaments contre le sida : alors que certains médicaments donnaient des résultats rapides et spectaculaires, certains médecins souhaitaient s'en tenir à une stricte application du protocole alors que des associations de malades jugeaient les preuves d'efficacité suffisantes et invitaient à diffuser sans plus

attendre le médicament, alors que la maladie se diffusait extrêmement vite [Dodier & Barbot 2000].

La reconnaissance implicite d'une pluralité de preuves acceptables (plus que d'une hiérarchie) doit être également pensée en économie. Il en va ainsi, par exemple, des politiques lancées en Chine qui ont abouti à un incontestable succès sur le plan macroéconomique [Rodrik 2008]. Elles ne procèdent pas d'expérimentations aléatoires qui auraient permis de trouver la bonne marche à suivre, et pourtant leur efficacité est incontestable. Certes, il demeure délicat d'isoler les différentes mesures ayant donné des résultats positifs et donc de mettre au jour les chaînes causales, mais l'efficacité des politiques macroéconomiques apparaît largement reconnue, sans qu'il y ait eu de recours à la randomisation. Pourtant, les randomistes tendent à affirmer qu'une politique économique qui ne serait pas évaluée *scientifiquement* – c'est-à-dire à les lire, en toute rigueur, évaluée par un essai randomisé contrôlé –, comme le suggèrent les économistes Marc Ferracci & Étienne Wasmer [Ferracci & Wasmer 2011], n'aurait pas lieu d'être.

On peut aller plus loin dans le parallèle avec la médecine. Certains acteurs du « mouvement randomiste » (le plus souvent les assistants de recherche, c'est-à-dire précisément ceux qui sont les plus proches de la *pratique* expérimentale, à l'instar des médecins pratiquant les essais, les économistes pratiquant des expérimentations aléatoires se rapprochant quant à eux davantage des biostatisticiens faisant essentiellement un « travail de bureau ») reconnaissent eux-mêmes, implicitement ou explicitement, le caractère flottant de cette hiérarchie qui placerait la randomisation tout en haut de l'évaluation [Jatteau 2016]. En effet, lorsque l'on écoute leurs discours, on se rend compte qu'*en pratique*, loin des discours officiels et de ce que l'on pourrait qualifier de « doxa randomiste », certains nuancent la supériorité de cette méthode. Pourtant, dans la littérature comme dans l'orientation de la recherche et des budgets, force est de constater qu'une seule méthode semble avoir droit de cité.

D'autres exemples de questionnements en médecine peuvent enrichir les pratiques expérimentales en économie. C'est le cas de ce que l'on appelle les « *horror stories* », c'est-à-dire les « histoires » qui montrent que pendant longtemps, en s'appuyant sur des études non randomisées, on a cru à l'efficacité d'un médicament jusqu'au jour où un essai clinique randomisé a montré qu'il était inefficace, voire, dans le pire des cas, qu'il était néfaste. Or ces études souffrent d'un certain nombre de biais : tentation de publier les *horror stories* uniquement si elles montrent que les essais cliniques randomisés obtiennent des résultats différents de ceux d'essais non randomisés inclus dans l'étude (dans certains cas, d'ailleurs, les premiers montrent que le traitement est efficace quand les seconds le voyaient inefficace [Rothwell 2005], chances de publication plus importantes pour les essais randomisés montrant l'inefficacité d'un traitement que pour les essais non randomisés montrant également l'inefficacité d'un traitement [Abel & Koch 1999]).

Dans de nombreux articles relatant des expérimentations aléatoires en économie, on retrouve cette idée de l'*horror story*, dans une version nuancée avec des conséquences moins dramatiques qu'en médecine. Enfin, avec la randomisation, on allait pouvoir obtenir des réponses définitives sur certains sujets, éventuellement en contradiction avec ce que l'on croyait auparavant. C'était ainsi l'ambition de Jessica Cohen & Pascaline Dupas, lorsqu'elles ont lancé une expérimentation aléatoire afin de savoir s'il fallait distribuer gratuitement des moustiquaires ou les vendre à prix subventionné [Cohen & Dupas 2010]. Avec la randomisation, elles souhaitaient apporter un argument d'un niveau de preuve jamais atteint jusqu'alors en matière de tarification des biens de santé.

Or la comparaison entre les essais randomisés et les essais non randomisés n'aboutit pas systématiquement à montrer que les résultats diffèrent et que les seconds obtiennent un effet plus important que les premiers, comme on le lit souvent dans la littérature médicale⁴. Nous pouvons prendre l'exemple d'une étude qui va précisément chercher à vérifier ces assertions en comparant sur cinq maladies différentes (tuberculose, cancer du sein, cholestérol, hypertension...) près d'une centaine d'articles relatant des essais cliniques randomisés et des études d'observation de cohortes ou de cas contrôlés [Concato, Shah *et al.* 2000]. Les chercheurs montrent ainsi que les résultats issus de méthodes non randomisées (mais « *well-designed* ») ne diffèrent pas significativement de la méthode randomisée. Leur démonstration va même plus loin, puisqu'elle montre que les études observationnelles ont une variance moins grande dans leurs résultats, comme on le voit sur le graphique ci-dessous de la figure 1. Une hypothèse pourrait être que ces dernières sont davantage financées par des fonds publics et ainsi moins soumises à la nécessité de trouver des résultats positifs, à l'inverse des essais cliniques randomisés massivement financés par les laboratoires privés et dont la recherche de résultats obéit à des impératifs économiques.

Sur le graphique, Les points noirs représentent les études randomisées, les points blancs les études observationnelles. On voit aisément que la dispersion des premiers est bien supérieure à celle des seconds. Autrement dit, avec les essais randomisés, on a plus de chance de trouver un chiffre éloigné de ce que l'on pense être le « vrai » résultat, en admettant que le « vrai » résultat se situe au niveau du point moyen. Sur l'axe des abscisses figurent les risques relatifs. Lorsque l'on se situe à droite de la limite pointillée (avec un risque relatif supérieur à 1), cela signifie que le groupe de traitement est davantage sujet à la maladie que le groupe de contrôle. Si la majorité des essais se situe bien à gauche, montrant ainsi l'efficacité du traitement, et laissant à penser que le « vrai » résultat se situe de ce côté, on remarque qu'à droite, on trouve

4. Comme le trouvent par exemple [Ioannidis, Haidich *et al.* 2001]. Ils montrent néanmoins que les effets vont globalement dans le même sens, que les essais soient randomisés ou non. C'est leur ampleur qui change. Il existe également des cas où les essais cliniques randomisés trouvent des effets de plus grande ampleur que les études non-randomisées [Boruch, Snyder *et al.* 2000].

Figure 1. Range of Point Estimates According to Type of Research Design.

The studies evaluated bacille Calmette–Guérin vaccine and active tuberculosis (13 randomized, controlled trials and 10 case–control studies), screening mammography and mortality from breast cancer (8 randomized, controlled trials and 4 case–control studies), cholesterol levels and death due to trauma among men (4 of 6 randomized, controlled trials [2 trials did not provide point estimates]; the results of the 14 cohort studies were not reported individually), treatment of hypertension and stroke among only the men in the studies (11 randomized, controlled trials and 7 cohort studies), and treatment of hypertension and coronary heart disease among only the men in the studies (13 of 14 randomized, controlled trials [1 trial did not provide point estimates] and 9 cohort studies). Solid circles indicate randomized, controlled trials, and open circles observational studies.

FIGURE 1 – Source de l'image [Concato, Shah *et al.* 2000], reproduction avec l'aimable autorisation de la MMS – Massachusetts Medical Society.

uniquement des essais randomisés. Comme le notent les auteurs, cela signifie que ce sont les seuls à aboutir à un résultat de sens opposé à celui du point moyen (c'est-à-dire de la majorité des essais) : les seuls à se tromper si l'on admet que la majorité des essais identifient le bon signe du « vrai » effet⁵. Les études non randomisées trouvent quant à elles toujours le bon « sens » de l'effet, et avec une variance bien moindre. Elles produisent donc des résultats à la fois plus justes et plus précis.

Que peut-on en retenir ? Que la supériorité des essais cliniques randomisés en matière d'évaluation n'est pas si unanimement reconnue et que le débat existe en médecine. Les randomistes, convaincus des bienfaits de la méthode qu'ils ont adoptée, semblent réfuter *a priori* les questionnements sur la hiérarchisation des preuves. Certes, la façon d'utiliser la randomisation en économie diffère partiellement de celle utilisée en médecine (absence d'usage de placebos, notamment⁶), mais les points communs sont bien plus nombreux.

5. Précisons que d'autres études trouvent à l'inverse que la variance des essais randomisés est moindre que les essais non randomisés [Kunz & Oxman 1998], [Benson & Hartz 2000]. La littérature ne montre ainsi pas une direction univoque en matière de comparaison entre les deux.

6. On compte cependant au moins une exception [Bulte, Pan *et al.* 2012].

Aussi ne peut-on que s'étonner que ces débats autour de la hiérarchie des preuves en médecine ne trouvent pas davantage d'écho chez les randomistes.

En conclusion provisoire, on peut dire que si la hiérarchie des preuves, et à travers elle la suprématie des essais cliniques randomisés, ne souffre pas de remise en cause généralisée en médecine, des questionnements existent, assortis d'arguments sérieux en faveur de l'utilisation d'une pluralité de modes d'administration de la preuve. La randomisation, malgré ses qualités qui ne doivent pas être niées, ne saurait être la solution idéale à tout problème d'évaluation, et quand elle en constitue une, elle n'est pas toujours la seule. C'est un enseignement clair que l'on peut tirer de la littérature médicale, et qui doit nous éclairer sur l'usage qui est fait de la méthode en économie.

2 Le problème de la généralisation des résultats

2.1 La validité externe, de la médecine à l'économie

Il existe un autre problème central en matière d'expérimentation : la généralisation des résultats, connue sous le terme de « validité externe » dans la littérature, par opposition à la « validité interne », qui renvoie à la certitude que l'on peut avoir que l'effet mesuré provient bien du traitement testé. Il est en effet fondamental de savoir si les effets procurés par un traitement sont susceptibles de se reproduire dans un autre contexte (temps, époque, population concernée). Par exemple, on se demande ce que l'administration d'un médicament est susceptible de produire dans un autre pays que celui où il a été testé, à une autre période et sur une population différente, ou alors dans le même pays mais sur une autre population, ou encore une population plus large. L'importance de la transposition d'un tel questionnement en économie est évidente.

Si la supériorité des expérimentations aléatoires n'est guère remise en cause par les randomistes, le problème de la validité externe est quant à lui bien documenté [Duflo & Kremer 2005], [Duflo 2004], [Duflo, Glennerster *et al.* 2008], [Banerjee & Duflo 2009], [Glennerster & Takavarasha 2013]. Mais il apparaît que ces économistes ne semblent pas prendre la mesure de cette potentielle limite de leur méthode. Ils considèrent – à juste titre – qu'il ne s'agit pas d'un problème spécifique aux expérimentations aléatoires⁷. Cela n'enlève pourtant rien de sa centralité, pour qui prétend avoir un impact sur les politiques publiques et la lutte contre la pauvreté – ce qui est bien l'objectif affiché des randomistes.

7. On peut juger que les essais cliniques randomisés comme les expérimentations aléatoires en économie ont toutefois tendance à l'exacerber, dans la mesure où ils se placent dans un cadre particulier, construit, à la différence de méthodes relevant davantage de l'observation qui ne s'inscrivent pas dans une logique expérimentale.

La validité externe est un problème bien connu en médecine. Cette question se pose depuis longtemps dans les recherches médicales, aussi bien dans les colonies au XVIII^e siècle [Chamayou 2013] que dans les premiers essais cliniques randomisés [Rothwell 2005]. Mais si le problème est documenté, il demeure pour certains négligé en médecine [Glasgow, Lichtenstein *et al.* 2003], [Rothwell 2006], voire sacrifié au profit de la validité interne [Solomon, Cavanaugh *et al.* 2009]. Pourtant, dans une perspective clinique, le fait de savoir si un résultat est transposable est une question de première importance pour les médecins, d'autant que certains effets cliniques comportent des biais importants (selon que le patient est une femme, un enfant...). L'effet bénéfique d'un traitement mesuré par un essai clinique randomisé est-il suffisant pour qu'un praticien l'applique directement sur un patient ?

Dans la pratique médicale, la question de la transposabilité clinique de la recherche est tout à fait primordiale [Glasgow, Green *et al.* 2006]. En effet, il ne s'agit pas seulement d'obtenir des résultats positifs des traitements testés, encore faut-il que ceux-ci soient applicables et appliqués hors du cadre dans lequel ils ont été obtenus [Rothwell 2005]. Or les essais cliniques randomisés manquent souvent de pertinence dans une optique clinique [Jacobson, Edwards *et al.* 1997]. Les recherches visant à prouver l'efficacité d'un traitement (preuve d'efficacité) sont bien plus nombreuses que les recherches cherchant à cerner l'effectivité d'un traitement (dans quelle mesure peut-il *effectivement* être utilisé dans la clinique et plus généralement sur une population différente de celle de l'essai ?). Cela revient à privilégier la validité interne au détriment de la validité externe [Black 1996], comme les grilles de lecture reconnues dans la profession le montrent, en privilégiant la première sur la seconde.

Si la traduction « sur le terrain » des résultats de la recherche est si importante, c'est que dans sa clinique, un médecin ne considère pas un « individu moyen », mais bien un patient particulier, avec son histoire médicale, ses pathologies prises dans leur globalité, ses attentes... Dès lors, les essais cliniques randomisés ne pouvant fournir des résultats qu'en moyenne (par la comparaison de moyennes entre les groupes), leur transposabilité à la clinique apparaît problématique. Il est tout à fait possible que des résultats globalement positifs cachent de fortes disparités entre les individus participant à une expérimentation (on parle d'hétérogénéité des effets). On peut imaginer un médicament qui permettrait de soigner une majeure partie d'une population, mais serait létal pour une petite minorité, par exemple. Si « en moyenne » il marche, peut-on néanmoins y avoir recours, sans connaître précisément la distribution de ces effets ? Appliquer des preuves apportées par les essais cliniques randomisés à des cas particuliers doit donc être fait avec précaution [Kravitz, Duan *et al.* 2004], comme le percevait déjà Claude Bernard [Supiot 2015].

Le parallèle avec l'économie est ici fondamental. Si le problème y est connu, il n'en reste pas moins que sa documentation en médecine peut être tout à fait utile. La question de l'applicabilité des résultats obtenus en économie est centrale, dans la mesure où les commanditaires de ces expérimentations (bien

souvent des organismes gouvernementaux ou internationaux, des fondations pour le développement, etc.) visent précisément une généralisation à terme des éventuels bienfaits montrés à l'aide d'une randomisation. Les acteurs locaux du développement cherchent avant tout des résultats opérationnels [Quentin & Guérin 2013]. Les questionnements posés en médecine à propos du transfert dans la pratique des résultats démontrés par les essais cliniques randomisés se posent donc avec acuité en économie. Cette facette de la validité externe conduit à s'interroger sur les conditions d'application d'un résultat à un autre contexte. De plus, se limiter à comparer des moyennes entre groupes peut être très insatisfaisant en termes de politiques publiques, car les gouvernants sont nécessairement intéressés autant par l'effet moyen d'une mesure que par les distributions de l'effet. Alors que les inégalités économiques s'avèrent croissantes dans les pays riches comme dans les pays pauvres, la prise en compte de l'hétérogénéité des effets apparaît incontournable [Gonzalez Alvaredo, Chancel *et al.* 2018].

2.2 Les contours de la validité externe

En médecine, la validité externe peut être abordée selon plusieurs points de vue, que l'on se place du côté de l'équipe médicale, des patients, du traitement lui-même ou en fonction du contexte [Rothwell 2005].

Prenons le cas des équipes médicales qui mettent en place les essais cliniques [Black 1996]. Les cliniciens qui participent à l'essai clinique peuvent ne pas être représentatifs de l'ensemble de ceux qui appliqueront ensuite éventuellement le traitement testé (c'est particulièrement vrai en chirurgie, mais pas seulement), par exemple, comme cela arrive, si seuls les meilleurs sont sélectionnés. Cela peut biaiser positivement les résultats qui, en cas de généralisation, risqueraient d'être moins bons dans la mesure où les praticiens seraient d'un niveau moindre que celui de ceux ayant participé à l'étude.

En économie, on retrouve des inquiétudes similaires. Les équipes chargées de mettre en place les expérimentations aléatoires bénéficient généralement d'une certaine expérience et peuvent être particulièrement enthousiastes, souhaitant que « ça marche », là où celles éventuellement chargées de généraliser ensuite peuvent avoir un comportement différent. On peut évoquer le cas d'une étude qui vise à voir si l'extension à tout un pays d'une expérience randomisée produit les mêmes effets [Bold, Kimenyi *et al.* 2013]. Une expérimentation aléatoire avait en effet montré, dans une petite région du Kenya, que l'embauche d'enseignants contractuels (et non fonctionnaires) améliorait les résultats des élèves aux tests scolaires [Duflo, Dupas *et al.* 2015]. Lors de l'expérimentation, c'était une ONG qui était chargée d'appliquer le programme. Lorsqu'on l'a étendu, le ministère kényan de l'Éducation en a été chargé et l'on s'est aperçu que les effets positifs, présents localement dans la phase expérimentale, disparaissaient lors de cette phase de généralisation. La résistance des syndicats des enseignants fonctionnaires à

cette extension à l'échelle nationale du programme, ainsi que la moindre implication du personnel du ministère de l'Éducation, sont avancées comme autant de pistes pour expliquer l'inefficacité du « traitement » une fois étendu. Ainsi, la généralisation d'un programme en dehors du contexte où il a été élaboré ne garantit pas la permanence des effets mesurés dans un certain contexte. Or, à l'heure actuelle en économie, on ne compte que très peu d'études comme celle-ci sur les effets de la généralisation d'un programme à une plus grande échelle.

La sélection des patients des essais, à l'instar de celle des cliniciens, doit également être questionnée du point de vue de la validité externe. S'ils diffèrent trop de la population cible (celle pour laquelle on pense que le traitement peut être bénéfique), alors la généralisation des résultats risque de poser problème [Kravitz, Duan *et al.* 2004]. Les critères de participation aux essais sont parfois si précis qu'ils excluent un très grand nombre de patients et que ceux retenus pour l'étude diffèrent significativement de ceux n'y participant pas [Batifoulier 2015]. De nombreuses études excluent les femmes et les personnes âgées, tandis que d'autres privilégient les patients qui sont les plus susceptibles de répondre favorablement au traitement.

On peut prendre l'exemple de deux essais menés sur des thérapies antirétrovirales [Moore, Goodall *et al.* 2000]. Les auteurs de l'article s'intéressent à leur potentiel de généralisation. La méthode qu'ils adoptent consiste à comparer les participants à l'étude à des malades qui y sont éligibles, mais qui n'y ont finalement pas participé. Cela revient à se demander si l'on peut généraliser les résultats à l'ensemble de la population d'intérêt. Or, les auteurs montrent que ces deux groupes diffèrent significativement selon des caractéristiques qu'ils jugent importantes (proportion d'homosexuels ou de toxicomanes, variété des pays de naissance...).

Une fois de plus, le parallèle avec l'économie est intéressant. Comment prétendre généraliser à une population qui diffère de celle à laquelle le programme testé est destiné ? La question de la représentativité est peu traitée par les randomistes, même si elle ne se pose pas toujours avec la même acuité : la représentativité d'une étude multicentrique menée en France dans différentes régions, comme celle sur la subvention des permis de conduire [L'Horty, Duguet *et al.* 2012], paraît meilleure qu'une expérimentation conduite dans une petite partie d'une région à l'ouest du Kenya [Glewwe, Kremer *et al.* 2009].

Les caractéristiques elles-mêmes du traitement peuvent limiter sa validité externe, par exemple lorsqu'elles sont trop originales par rapport à ce qu'il est possible de faire dans des conditions normales [Black 1996] ou lorsque le protocole utilisé diffère trop de ceux habituellement suivis [Rothwell 2005].

Le choix des « *outcomes* », ces variables qui permettent de mesurer les effets, peut également poser problème. Dans bon nombre d'essais cliniques randomisés, les *outcomes* retenus sont des « *surrogate outcomes* », autrement dit des *outcomes* de substitution ou *outcomes* intermédiaires (utilisés quand il n'est pas possible d'en obtenir de meilleurs). Or, ils peuvent être trompeurs quant

à l'impact sur les véritables variables d'intérêt [Labrousse 2010]. Envisager de généraliser à partir de ces *outcomes* peut alors avoir des effets néfastes. De plus, les *outcomes* retenus par les chercheurs ne sont pas toujours ceux qui intéressent le plus les praticiens, ni même les patients, en médecine comme en économie, particulièrement dans le monde du développement [Quentin & Guérin 2013].

Les problèmes que ces différents points soulèvent se retrouvent en économie. Pour certaines expérimentations aléatoires, le traitement apparaît trop original pour envisager une généralisation. L'expérimentation visant à limiter l'absentéisme des instituteurs en Inde à l'aide d'incitations financières, leur présence en classe étant contrôlée à l'aide de photographies qu'ils doivent prendre de leurs élèves et d'eux-mêmes à différents moments de la journée en est un bon exemple [Duflo, Hanna *et al.* 2012]. Bien que l'article montre que les incitations « marchent », on peut douter d'une généralisation qui verrait le gouvernement indien acheter des millions d'appareils photographiques... Ce même exemple peut être utile pour illustrer la difficulté à choisir les « bons » *outcomes*. Certes, le taux de présence s'accroît, mais à quel prix en termes d'estime de soi et de motivation intrinsèque ? D'autres *outcomes* auraient peut-être pu donner un sens différent aux conclusions avancées par les auteurs.

Enfin, il convient de prendre en compte également le contexte dans lequel se déroulent les essais cliniques randomisés, celui-ci étant à même d'avoir un effet qui altère la validité externe [Solomon, Cavanaugh *et al.* 2009]. Le cadre dans lequel a lieu un essai clinique n'est certainement pas neutre et généraliser un traitement dans un autre contexte peut modifier les effets qu'il produit. Ainsi, les différences entre les systèmes de santé peuvent mettre à mal la validité externe. Un médicament dont un essai aura montré qu'il est efficace contre une maladie ne sera pas nécessairement consommé par les patients, en fonction de son prix, de sa disponibilité et plus généralement de la présence et de la forme d'une sécurité sociale. La localisation est donc à même d'avoir un impact sur les effets mesurés et une généralisation à un autre système de santé ne garantit pas des résultats similaires. Même pour des systèmes de santé proches, les différences entre les pays peuvent miner la validité externe. Là encore, dans une perspective de lutte contre la pauvreté, la question du contexte dans les expérimentations aléatoires est tout à fait primordiale afin d'essayer d'approcher au mieux les *réels* effets d'un « traitement ».

2.3 Dépasser le problème

Toutes les limites quant à la validité externe que nous avons décrites amènent à considérer les pistes qui peuvent être suivies pour répondre à ce problème. Il ne s'agit pas tant de rejeter les essais contrôlés randomisés parce que leurs résultats ne sont pas valables pour tous, en tout temps et en tout lieu, que de préciser pour qui, quand et dans quel contexte ces résultats sont valables. L'objectif n'est ainsi pas d'avoir nécessairement une forte validité externe, mais

de pouvoir juger de la généralisation potentielle du traitement. Un résultat qui ne serait que peu généralisable n'est pas inintéressant, pour peu qu'on soit conscient de sa faible validité externe.

Dans cette optique, certains auteurs proposent de chercher des preuves d'applicabilité des solutions testées, à l'aide d'études non randomisées, en cherchant les effets non pas pour un groupe de patients particulier, mais bien pour la population susceptible de bénéficier des traitements dans son ensemble [Kravitz, Duan *et al.* 2004]. Pour espérer de l'*Evidence-Based Practice*, il convient de faire de la *Practice-Based Evidence* [Glasgow, Green *et al.* 2006]. La distinction entre efficacité et effectivité, la première regardant uniquement s'il y a un effet, quand la deuxième se demande s'il peut y avoir un effet dans les conditions réelles [Barrett & Carter 2014], qui est bien établie en médecine, devrait inspirer les économistes.

L'hétérogénéité géographique des effets peut mettre à mal la validité externe [Horwitz, Singer *et al.* 1996]. Elle peut provenir d'une disparité géographique des centres où sont pratiqués les essais. En médecine, on préconise ainsi souvent des essais dans différents lieux. Cela constitue souvent une condition requise par les revues et/ou les financeurs. Elle est aussi très ancienne, puisque l'on trouve des essais multicentriques dès 1934 [Isambert 1987]. Le rôle des éditeurs de revue doit également être mentionné. Plusieurs militent pour une meilleure prise en compte dans les journaux scientifiques du problème de la validité externe, invitant ainsi à toujours vérifier que la question « à qui ces résultats s'appliquent-ils ? » est bien posée dans les articles qu'ils publient [Rothwell 2005].

Les recommandations formulées à propos des essais cliniques randomisés en médecine sont pour la plupart transposables à l'économie. Un principe de précaution doit tout d'abord se faire jour : il s'agit de ne jamais laisser entendre que les résultats obtenus à l'aide d'une expérimentation aléatoire possèdent *a priori* une validité externe. Or nombreux sont les articles les relatant où l'on peut constater des glissements à ce sujet [Jatteau 2011]. Une preuve d'efficacité ne vaut pas preuve de généralisation. La rigueur scientifique impose la circonscription des résultats à leur contexte et invite à donner des éléments qui permettent de juger la validité externe. Trop souvent, la généralisation est laissée à l'appréciation des chercheurs qui, notamment dans leurs conclusions, peuvent avoir tendance à exagérer la validité externe. Il ne s'agit pas de dénier toute valeur à une expérimentation aléatoire si elle n'a pas de validité externe forte, c'est-à-dire qu'on ne peut généraliser le « traitement » sans être certain que les effets aillent toujours dans le sens mis en avant par l'expérimentation. Mais elle n'a que peu d'intérêt si elle n'informe pas sur les contours possibles d'une généralisation. Il ne faut pas négliger l'opérationnalité des résultats obtenus par des expérimentations aléatoires. S'ils devaient rester cantonnés au monde de la recherche, leur utilité en serait amoindrie, en particulier aux yeux des financeurs de ces études, que ce soit des agences d'aide au développement, des fondations privées ou des gouvernements. L'applicabilité des résultats doit être pensée à hauteur de l'importance que lui accordent les parties prenantes

autres que les chercheurs, aussi bien au niveau du *design* de l'expérimentation que des résultats [Quentin & Guérin 2013]. Il faut dire un mot des individus qui sont engagés dans ces expérimentations, qu'ils soient dans le groupe de contrôle ou dans celui de traitement. Paradoxalement, ils apparaissent assez absents des essais contrôlés randomisés et ils ne sont approchés qu'à travers les données quantitatives collectées à partir des questionnaires auxquels ils répondent. Or c'est bien à travers eux d'abord que doit être saisie l'importance de la validité externe, dans la mesure où ils sont potentiellement les premiers bénéficiaires d'une telle extension, au-delà des retombées académiques en termes de publications dont peuvent bénéficier les chercheurs qui lancent ces essais et dont l'attente peut d'ailleurs entrer en conflit avec bienfaits pour les populations [Jatteau 2013].

Il s'agit de faire preuve de vigilance quant aux généralisations qui se font à partir d'une seule étude, comme c'est le cas en médecine, d'autant que les conditions de généralisation sont rarement pensées avec les expérimentations aléatoires. Pour mieux prendre en compte la validité externe, sans doute faut-il plus que de la randomisation, pour reprendre le titre d'un article à ce sujet [Glasgow, Green *et al.* 2006]. D'autres types d'études doivent être considérés, et la méfiance qui est celle des randomistes à l'égard des méthodes qualitatives doit être levée. C'est dans une pluralité de méthodes, plutôt que dans une stricte hiérarchie, que certains chercheurs sur les essais cliniques randomisés voient le salut en matière de validité externe. D'après eux, elles permettent une contextualisation plus fine, mieux à même de fournir des éléments pour permettre de juger d'une généralisation potentielle. Un tel pluralisme doit être considéré en économie.

3 Conclusion

Que le tirage au sort soit utilisé dans un cadre expérimental pour mesurer l'effet d'un traitement, en médecine comme en économie, apparaît satisfaisant d'un point de vue statistique. Théorisée par Fisher dans l'entre-deux guerres en s'appuyant sur des tentatives antérieures [Fisher 1925, 1935], la méthode de l'expérimentation aléatoire est depuis utilisée dans de nombreuses disciplines scientifiques. La médecine est certainement celle qui a contribué à la populariser, à travers les dizaines de milliers d'essais cliniques randomisés lancés depuis la seconde guerre mondiale [Olkin 1995]. À côté, l'économie fait pâle figure, même en prenant en compte, en plus de la vague actuelle représentée par l'économiste Esther Duflo, les expérimentations précoces (et méconnues) de l'entre-deux guerres [Barbier 2012], comme celles, davantage documentées, des années 1960-1980 aux États-Unis [Monnier 1992]. La première a emmagasiné bien plus d'expériences que la seconde, et les discussions théoriques sur la méthode des essais contrôlés randomisés n'en sont que plus riches. L'hypothèse de travail de cet article était ainsi que la médecine pouvait faire bénéficier

l'économie de son recul historique supérieur, dont découle une réflexion plus poussée. Deux éléments en particulier nous semblent devoir être mis en avant.

Le premier invite à nuancer le transfert de hiérarchisation des formes de preuves qu'ont opéré les randomistes. S'appuyant implicitement sur la hiérarchie des preuves telle qu'elle est diffusée en médecine, ils arguent de la supériorité de la randomisation sur toute autre forme de méthode d'évaluation. Si on retrouve un tel classement dans la littérature médicale, celle-ci comporte également tout un pan de recherches visant à questionner une telle hiérarchie. Il apparaît ainsi que *même* en médecine, le tirage au sort comme outil ultime pour mesurer l'effet d'un traitement est régulièrement remis en cause. Cela implique corollairement une (re)valorisation des autres méthodes. Ce relatif pluralisme méthodologique revendiqué par des chercheurs en médecine, et plus encore les raisons qui les poussent à le défendre, doivent être considérés par les randomistes. Il s'agit de sortir d'une posture confortable qui laisserait entendre que la supériorité d'une méthode aurait été définitivement prouvée, en se référant à une discipline (la médecine) dont l'aura scientifique, si elle peut être questionnée, est certainement supérieure à celle de l'économie.

La question de la généralisation se pose différemment, car ce problème n'est pas nié par les randomistes, mais il demeure néanmoins dévalorisé. Sur cet aspect également, la littérature médicale permet d'éclairer à quel point elle est cruciale dans une optique d'application des résultats. Pour sortir de l'expérimentation et aller vers la généralisation, il faut considérer la validité externe à sa juste valeur. Cette question est ancienne en médecine, où les chercheurs se sont depuis longtemps saisis de cette question. Outre que leurs travaux permettent d'en saisir toute l'importante en économie, les pistes qu'ils proposent pour le dépasser doivent également être prises en compte par les randomistes, notamment en ce qu'elles appellent à une diversité méthodologique accrue.

C'est là l'enseignement principal de cette comparaison entre la médecine et l'économie à travers le prisme des essais contrôlés randomisés. Quoique puissant statistiquement, ils apparaissent insuffisants à eux seuls pour remplir l'objectif qu'on leur assigne : non seulement montrer les effets d'un traitement dans un cadre expérimental, mais aussi et surtout, montrer quels effets un tel traitement aurait une fois généralisé.

Bibliographie

- ABEL, Ulrich & KOCH, Armin [1999], The role of randomization in clinical studies : Myths and beliefs, *Journal of Clinical Epidemiology*, 52(6), 487–497, doi : 10.1016/S0895-4356(99)00041-4.
- ABRAHAM, John & REED, Tim [2002], Progress, innovation and regulatory science in drug development : The politics of international standard-setting, *Social Studies of Science*, 32(3), 337–369, doi : 10.2307/3183031.

- ARMITAGE, Peter [2003], Fisher, Bradford Hill, and randomization, *International Journal of Epidemiology*, 32(6), 925–928, doi : 10.1093/ije/dyg286.
- ASHRAF, Nava, BERRY, James *et al.* [2010], Can higher prices stimulate product use? Evidence from a field experiment in Zambia, *American Economic Review*, 100(5), 2383–2413, doi : 10.1257/aer.100.5.2383.
- BANERJEE, Abhijit V. (éd.) [2007], *Making Aid Work*, Cambridge : The MIT Press.
- BANERJEE, Abhijit V. & DUFLO, Esther [2009], L’approche expérimentale en économie du développement, *Revue d’Économie politique*, 119, 691–726, doi : 10.3917/redp.195.0691.
- BARBIER, Jean-Claude [2012], évaluations “expérimentales” ? : quelques leçons de l’histoire, *Politiques sociales et familiales*, 110, 19–31, doi : 10.3406/cf.2012.2731.
- BARRETT, Christopher B. & CARTER, Michael R. [2014], Retreat from radical skepticism : Rebalancing theory, observational data and randomization in development economics, dans : *Fields Experiments and Their Critics*, édité par D. L. Teele, New Haven ; Londres : Yale University Press, 58–77.
- BATIFOULIER, Philippe [2015], Aux origines de la privatisation du financement du soin : quand la théorie de l’aléa moral rencontre le capitalisme sanitaire, *Revue de la régulation*, 17, 2–32, doi : 10.4000/regulation.11196.
- BENSON, Kjell & HARTZ, Arthur J. [2000], A comparison of observational studies and randomized, controlled trials, *New England Journal of Medicine*, 342(25), 1878–1886, doi : 10.1056/NEJM200006223422506.
- BLACK, Nick [1996], Why we need observational studies to evaluate the effectiveness of health care, *British Medical Journal*, 312(7040), 1215–1218, doi : 10.1136/bmj.312.7040.1215.
- BOLD, Tessa, KIMENYI, Mwangi *et al.* [2013], Scaling up what works : Experimental evidence on external validity in Kenyan education, Working Paper, Rap. tech., Center for Global Development.
- BORUCH, Robert, SNYDER, Brooke *et al.* [2000], The importance of randomized field trials, *Crime & Delinquency*, 46(2), 156–180, doi : 10.1177/0011128700046002002.
- BRUHN, Miriam & MCKENZIE, David [2009], In pursuit of balance : Randomization in practice in development field experiments, *American Economic Journal : Applied Economics*, 1(4), 200–232, doi : 10.2307/25760187.

- BULTE, Erwin, PAN, Lei *et al.* [2012], Pseudo-placebo effects in randomized controlled trials for development : Evidence from a double-blind field experiment in Tanzania, Working Paper.
- CARTWRIGHT, Nancy [2007], Are RCTs the gold standard ?, *BioSocieties*, 2(1), 11–20, doi : 10.1017/S1745855207005029.
- CHAMAYOU, Grégoire [2013], *Les Corps vils. Expérimenter sur les êtres humains aux XVIII^e et XIX^e siècles*, Paris : La Découverte.
- CHARLTON, B. G. & MILES, A. [1998], The rise and fall of EBM, *QJM : An International Journal of Medicine*, 91(5), 371–374, doi : 10.1093/qjmed/91.5.371.
- COHEN, Jessica & DUPAS, Pascaline [2010], Free distribution or cost-sharing ? evidence from a randomized malaria prevention experiment, *The Quarterly Journal of Economics*, 125(1), 1–45, doi : 10.1162/qjec.2010.125.1.1.
- CONCATO, John, SHAH, Nirav *et al.* [2000], Randomized, controlled trials, observational studies, and the hierarchy of research designs, *New England Journal of Medicine*, 342(25), 1887–1892, doi : 10.1056/NEJM200006223422507.
- CORIAT, Benjamin, COUTROT, Thomas *et al.* (éds.) [2017], *Misère du scientisme en économie. À propos de l'affaire Cahuc et Zylberberg*, Vulaines-sur-Seine : Éditions du Croquant.
- DEATON, Angus [2010], Instruments, randomization, and learning about development, *Journal of Economic Literature*, 48, 424–455, doi : 10.1056/NEJM200006223422506.
- DODIER, Nicolas & BARBOT, Janine [2000], Le temps des tensions épistémiques : Le développement des essais thérapeutiques dans le cadre du sida, *Revue française de sociologie*, 41(1), 79–118, doi : 10.2307/3322655.
- DUFLO, Esther [2004], Scaling up and evaluation, dans : *Accelerating Development*, édité par F. Bourguignon & B. Pleskovic, Washington, DC ; Oxford : World Bank ; Oxford University Press, 342–367.
- [2010a], *La Politique de l'autonomie*, Paris : Seuil.
- [2010b], *Le Développement humain*, Paris : Seuil.
- DUFLO, Esther, DUPAS, Pascaline *et al.* [2015], School governance, teacher incentives, and pupil–teacher ratios : Experimental evidence from Kenyan primary schools, *Journal of Public Economics*, 123, 92–110, doi : 10.1016/j.jpubeco.2014.11.008.

- DUFLO, Esther, GLENNERSTER, Rachel *et al.* [2008], Using randomization in development economics research : A toolkit, dans : *Handbook of Development Economics*, édité par T. P. Schultz & J. A. Strauss, Burlington : Elsevier, t. 4, 3895–3962, doi : 10.1016/S1573-4471(07)04061-2.
- DUFLO, Esther, HANNA, Rema *et al.* [2012], Incentives work : Getting teachers to come to school, *American Economic Review*, 102(4), 1241–1278, doi : 10.1257/aer.102.4.1241.
- DUFLO, Esther & KREMER, Michael [2005], Use of randomization in the evaluation of development effectiveness, dans : *Evaluating Development Effectiveness*, édité par G. K. Pitman, N. F. Osvaldo & G. K. Ingram, New Brunswick : Transaction Publishers, 93–120.
- DUFLO, Esther, KREMER, Michael *et al.* [2008], How high are rates of return to fertilizer ? Evidence from field experiments in Kenya, *American Economic Review*, 98(2), 482–88, doi : 10.1257/aer.98.2.482.
- EBER, Nicolas & WILLINGER, Marc [2012], *L'Économie expérimentale*, Paris : La Découverte.
- EBLE, Alex, BOONE, Peter *et al.* [2014], Risk and evidence of bias in randomized controlled trials in economics, Rap. tech., Brown University, Mimeo.
- FEINSTEIN, Alvan R. [1984], Current problems and future challenges in randomized clinical trials, *Circulation*, 70(5), 767–774, doi : 10.1161/01.CIR.70.5.767.
- FERRACCI, Marc & WASMER, Étienne [2011], *État moderne, État efficace. Evaluer les dépenses publiques pour sauvegarder le modèle français*, Paris : Odile Jacob.
- FISHER, Ronald A. [1925], *Statistical Methods for Research Workers*, Edinburgh : Oliver & Boyd.
- [1935], *Design of Experiments*, Edinburgh : Oliver & Boyd.
- GLASGOW, Russell E., GREEN, Lawrence W. *et al.* [2006], External validity : We need to do more, *Annals of Behavioral Medicine*, 31(2), 105–108, doi : 10.1207/s15324796abm3102_1.
- GLASGOW, Russell E., LICHTENSTEIN, Edward *et al.* [2003], Why don't we see more translation of health promotion research to practice ? Rethinking the efficacy-to-effectiveness transition, *American Journal of Public Health*, 93(8), 1261–1267, doi : 10.2105/AJPH.93.8.1261.
- GLENNERSTER, Rachel & TAKAVARASHA, Kuzai [2013], *Running Randomized Evaluations : A Practical Guide*, Princeton : Princeton University Press.

- GLEWWE, Paul, KREMER, Michael *et al.* [2009], Many children left behind? Textbooks and test scores in Kenya, *American Economic Journal : Applied Economics*, 1(1), 112–135, doi : 10.1257/app.1.1.112.
- GONZALEZ ALVAREDO, Facundo, CHANCEL, Lucas *et al.* (éds.) [2018], *Rapport sur les inégalités mondiales*, Paris : Seuil.
- GUERON, Judith M. [2002], The politics of random assignment : Implementing studies and affecting policy, dans : *Evidence Matters : Randomized Trials in Education Research*, édité par F. Mosteller & R. F. Boruch, Washington : Brookings Institution Press, 15–49.
- GUERON, Judith M. & ROLSTON, Howard [2013], *Fighting for Reliable Evidence*, New York : Russell Sage Foundation Publications.
- HENRY-AMAR, Michel [2005], Études de cohortes, essais cliniques randomisés, méta-analyse : quelle méthode pour quels résultats?, dans : *27^e journées de la SFSPM*, 40–47.
- HOLMES, Dave, MURRAY, Stuart J. *et al.* [2006], Deconstructing the evidence-based discourse in health sciences : Truth, power and fascism, *International Journal of Evidence-Based Healthcare*, 4(3), 180–186, doi : 10.1111/j.1479-6988.2006.00041.x.
- HORWITZ, Ralph I., SINGER, Burton H. *et al.* [1996], Can treatment that is helpful on average be harmful to some patients? A study of the conflicting information needs of clinical inquiry and drug regulation, *Journal of Clinical Epidemiology*, 49(4), 395–400, doi : 10.1016/0895-4356(95)00058-5.
- IOANNIDIS, John A., HAIDICH, Anna-Bettina *et al.* [2001], Comparison of evidence of treatment effects in randomized and nonrandomized studies, *Journal of the American Medical Association*, 286(7), 821–830, doi : 10.1001/jama.286.7.821.
- ISAMBERT, François-André [1987], L'expérimentation sur l'homme comme pratique et comme représentation, *Actes de la recherche en sciences sociales*, 68, 15–30.
- JACOBSON, Lionel D., EDWARDS, Adrian G. K. *et al.* [1997], Evidence-based medicine and general practice, *British Journal of General Practice*, 47(420), 449–452, <https://bjgp.org/content/47/420/449>.
- JATTEAU, Arthur [2011], *Les Expérimentations aléatoires en économie du développement : enjeux critiques, pratique de terrain et portée politique*, Mémoire de master 2, EHESS, Paris.
- [2013], Expérimenter le développement ? Des économistes et leur terrain, *Genèses*, 4(93), 8–28, doi : 10.3917/gen.093.0008.

- [2016], *Faire preuve par le chiffre ? Le cas des expérimentations aléatoires en économie*, Thèse de doctorat, ENS Paris Saclay.
- [2018], Comment expliquer le succès de la méthode des expérimentations aléatoires ? Une sociographie du J-PAL, *SociologieS*, 13 mars, en ligne, <https://journals.openedition.org/sociologies/6700#quotation>.
- KRAVITZ, Richard L., DUAN, Naihua *et al.* [2004], Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages, *The Milbank Quarterly*, 82(4), 661–687, doi : 10.1111/j.0887-378X.2004.00327.x.
- KUNZ, Regina & OXMAN, Andrew D. [1998], The unpredictability paradox : Review of empirical comparisons of randomised and non-randomised clinical trials, *British Medical Journal*, 317(7167), 1185–1190, doi : 10.1136/bmj.317.7167.1185.
- LABROUSSE, Agnès [2010], Nouvelle économie du développement et essais cliniques randomisés : une mise en perspective d’un outil de preuve et de gouvernement, *Revue de la régulation*, 7, 2–32, doi : 10.4000/regulation.7818.
- [2016], Apprendre des expérimentations aléatoires. Promesse de scientificité, complications pratiques, expériences historiques, *La Vie des idées*, 4 mars, 1–10, https://lavedesidees.fr/IMG/pdf/20160304_labrousse.pdf.
- LAURENT, Catherine, BAUDRY, Jacques *et al.* [2009], Pourquoi s’intéresser à la notion d’“evidence-based policy” ?, *Tiers Monde*, 200(4), 853–873, doi : 10.3917/rtm.200.0853.
- LEVITT, Steven D. & LIST, John A. [2009], Field experiments in economics : The past, the present, and the future, *European Economic Review*, 53(1), 1–18, doi : 10.1016/j.eurocorev.2008.12.001.
- L’HORTY, Yannick, DUGUET, Emmanuel *et al.* [2012], Faut-il subventionner le permis de conduire des jeunes en difficulté d’insertion ?, Rapport d’évaluation quantitative, Fonds d’expérimentation pour la jeunesse, s.l.
- L’HORTY, Yannick & PETIT, Pascale [2011], évaluation aléatoire et expérimentations sociales, *Revue française d’économie*, 26(4), 13–48, doi : 10.3917/rfe.111.0013.
- LORDON, Frédéric [1997], Le désir de “faire science”, dans : *Actes de la recherche en sciences sociales*, t. 119, 27–35.
- MARKS, Harry [1997], *The Progress of Experiment : science and therapeutic reform in the United States (1900–1990)*, Cambridge : Cambridge University Press, trad. fr. par F. Bouillot, *La Médecine des preuves : histoire et anthropologie des essais cliniques (1900–1990)*, Le Plessis-Robinson : Institut Synthélabo pour le progrès de la connaissance, 1999.

- MIGUEL, Edward & KREMER, Michael [2004], Worms : Identifying impacts on education and health in the presence of treatment externalities, *Econometrica*, 72(1), 159–217, doi : 10.1111/j.1468-0262.2004.00481.x.
- MONNIER, Éric [1992], *Évaluations de l'action des pouvoirs publics*, Paris : Economica.
- MOORE, D. A. J., GOODALL, R. I. *et al.* [2000], How generalizable are the results of large randomized controlled trials of antiretroviral therapy ?, *HIV Medicine*, 1(3), 149–154, doi : 10.1046/j.1468-1293.2000.00019.x.
- NAYLOR, C.D. [1995], Grey zones of clinical practice : some limits to evidence-based medicine, *The Lancet*, 345(8953), 840–842, doi : 10.1016/S0140-6736(95)92969-X.
- OLKIN, Ingram [1995], Statistical and theoretical considerations in meta-analysis, *Journal of Clinical Epidemiology*, 48(1), 133–146, doi : 10.1016/0895-4356(94)00136-E.
- PICCIOTTO, Robert [2012], Experimentalism and development evaluation : Will the bubble burst ?, *Evaluation*, 18(2), 213–229, doi : 10.1177/1356389012440915.
- QUENTIN, Aurélie & GUÉRIN, Isabelle [2013], La randomisation à l'épreuve du terrain. L'exemple du projet de microassurance Sky au Cambodge, *Revue Tiers Monde*, 213(1), 179–200, doi : 10.3917/rtm.213.0179.
- RODRIK, Dani [2008], The New Development Economics : We Shall Experiment, but How Shall We Learn ?, Working Paper 55, HKS.
- ROTHWELL, Peter M. [2005], External validity of randomised controlled trials : “to whom do the results of this trial apply”, *The Lancet*, 365(9453), 82–93, doi : 10.1016/S0140-6736(04)17670-8.
- [2006], Factors that can affect the external validity of randomised controlled trials, *PLOS Clinical Trials*, 1(1), 1–5, doi : 10.1371/journal.pctr.0010009.
- SCHULZ, Kenneth F. [1995], Subverting randomization in controlled trials, *JAMA*, 274(18), 1456–1458, doi : 10.1001/jama.1995.03530180050029.
- SCHULZ, Kenneth F. & GRIMES, David A. [2002], Blinding in randomised trials : hiding who got what, *The Lancet*, 359(9307), 696–700, doi : 10.1016/S0140-6736(02)07816-9.
- SCRIVEN, Michael [2008], A summative evaluation of rct methodology : An alternative approach to causal research, *Journal of MultiDisciplinary Evaluation*, 5(9), 11–24, http://journals.sfu.ca/jmde/index.php/jmde_1/article/view/160.

- SIGN 50 [2008], A guideline developer's handbook, Rap. tech., Scottish Intercollegiate Guidelines Network, s. l.
- SMITH, Gordon C. S. & PELL, Jill P. [2003], Parachute use to prevent death and major trauma related to gravitational challenge : Systematic review of randomised controlled trials, *British Medical Journal*, 327, 1459–1461, doi : 10.1136/bmj.327.7429.1459.
- SOLOMON, Phyllis L., CAVANAUGH, Mary M. *et al.* [2009], *Randomized Controlled Trials : Design and Implementation for Community-based Psychosocial Interventions*, New York : Oxford University Press.
- SUPIOT, Alain [2015], *La Gouvernance par les nombres. Cours au Collège de France (2012-2014)*, Paris : Fayard.
- The Lancet* [2004], The World Bank is finally embracing science, *The Lancet*, 364(9436), 731–732, doi : 10.1016/S0140-6736(04)16945-6.
- THOMAS, James, HARDEN, Angela *et al.* [2004], Integrating qualitative research with trials in systematic reviews, *British Medical Journal*, 328(7446), 1010–1012, doi : 10.1136/bmj.328.7446.1010.
- THORNTON, Rebecca L. [2008], The demand for, and impact of, learning HIV status, *American Economic Review*, 98(5), 1829–1863, doi : 10.1257/aer.98.5.1829.
- WORRALL, John [2007], Evidence in medicine and evidence-based medicine, *Philosophy Compass*, 2(6), 981–1022, doi : 10.1111/j.1747-9991.2007.00106.x.