

## Papiers de Recherche | Research Papers

---

# L'étalon-or des évaluations randomisées : Économie politique des expérimentations aléatoires dans le domaine du développement

---

Florent BÉDÉCARRATS  
Isabelle GUÉRIN  
François ROUBAUD

Mai 2017

**Pour citer ce papier :**

BÉDÉCARRATS F., I. GUÉRIN et F. ROUBAUD (2017), "L'étalon-or des évaluations randomisées : économie politique des expérimentations aléatoires dans le domaine du développement", *Papiers de Recherche AFD*, n° 2017-44, Mai.

**Contact à l'AFD :**

Florent BÉDÉCARRATS (bedecarratsf@afd.fr)

## Papiers de Recherche de l'AFD

Les *Papiers de Recherche de l'AFD* ont pour but de diffuser rapidement les résultats de travaux en cours. Ils s'adressent principalement aux chercheurs, aux étudiants et au monde académique. Ils couvrent l'ensemble des sujets de travail de l'AFD : analyse économique, théorie économique, analyse des politiques publiques, sciences de l'ingénieur, sociologie, géographie et anthropologie. Une publication dans les *Papiers de Recherche de l'AFD* n'en exclut aucune autre.

L'Agence Française de Développement (AFD), institution financière publique qui met en œuvre la politique définie par le gouvernement français, agit pour combattre la pauvreté et favoriser le développement durable. Présente sur quatre continents à travers un réseau de 72 bureaux, l'AFD finance et accompagne des projets qui améliorent les conditions de vie des populations, soutiennent la croissance économique et protègent la planète. En 2014, l'AFD a consacré 8,1 milliards d'euros au financement de projets dans les pays en développement et en faveur des Outre-mer.

**Les opinions exprimées dans ce papier sont celles de son (ses) auteur(s) et ne reflètent pas nécessairement celles de l'AFD. Ce document est publié sous l'entière responsabilité de son (ses) auteur(s).**

Les *Papiers de Recherche* sont téléchargeables sur : <http://librairie.afd.fr/>

## AFD Research Papers

*AFD Research Papers* are intended to rapidly disseminate findings of ongoing work and mainly target researchers, students and the wider academic community. They cover the full range of AFD work, including: economic analysis, economic theory, policy analysis, engineering sciences, sociology, geography and anthropology. *AFD Research Papers* and other publications are not mutually exclusive.

Agence Française de Développement (AFD), a public financial institution that implements the policy defined by the French Government, works to combat poverty and promote sustainable development. AFD operates on four continents via a network of 72 offices and finances and supports projects that improve living conditions for populations, boost economic growth and protect the planet. In 2014, AFD earmarked EUR 8.1bn to finance projects in developing countries and for overseas France.

**The opinions expressed in this paper are those of the author(s) and do not necessarily reflect the position of AFD. It is therefore published under the sole responsibility of its author(s).**

*AFD Research Papers* can be downloaded from: <http://librairie.afd.fr/en/>

AFD, 5 rue Roland Barthes

75598 Paris Cedex 12, France

✉ [ResearchPapers@afd.fr](mailto:ResearchPapers@afd.fr)

ISSN 2492 - 2846

## **L'étalon-or des évaluations randomisées : économie politique des expérimentations aléatoires dans le domaine du développement**

Florent Bédécarrats, AFD Evaluation Unit.

Isabelle Guérin, IRD-CESSMA.

François Roubaud, IRD-DIAL.

### **Résumé**

La dernière décennie a vu l'émergence d'un nouveau champ de recherche en économie du développement : les méthodes expérimentales d'évaluation d'impacts par assignation aléatoire. Cet article explore le contraste entre d'une part les limites (nombreuses) et la circonscription (très étroite) du champ réel d'application de ces méthodes et d'autre part leur succès, attesté à la fois par leur nombre et leur forte médiatisation. L'analyse suggère que ce contraste est le fruit d'une conjonction économique et politique particulière, émanant de stratégies novatrices de la part des chercheurs de cette nouvelle école, et d'intérêts et de préférences spécifiques provenant à la fois du monde académique et de la communauté bailleurs de fond.

**Mots-clés :** Evaluation d'impact, Méthode expérimentale, Essai randomisé, Méthodologie, Economie politique, Développement

**Classification JEL :** A11, B41, C18, C93, D72, O10.

**Version originale :** Français

**Acceptée :** Mai 2017

## Introduction

La dernière décennie a vu l'émergence d'un nouveau champ de recherche en économie du développement : les méthodes expérimentales d'évaluation d'impacts par assignation aléatoire (*Randomized Control Trial* ; dans la suite du texte RCT). Si le principe des RCT n'est pas une innovation scientifique, son application à grande échelle dans les pays en développement (PED) est en revanche un phénomène sans précédent. Dérivées des sciences médicales, elles sont déjà utilisées depuis les années 1960 pour l'évaluation des politiques publiques dans les pays développés (essentiellement aux États-Unis, dans les domaines comme la criminalité, l'assurance, l'emploi, la fiscalité ou encore l'éducation ; Oakley, 2000 ; Pritchett *et alii*, 2013). Ces méthodes ont ensuite été adaptées aux problématiques et aux contextes des pays pauvres. Elles ont connu un succès retentissant, ce qu'atteste leur multiplication. Leurs promoteurs ont rapidement bénéficié d'une véritable consécration internationale. Les meilleures revues d'économie ont ouvert grand leurs portes aux RCT, comme le montre le nombre sans cesse croissant d'articles publiés. Cette renommée dépasse aujourd'hui le champ disciplinaire de l'économie : jusqu'à la prestigieuse revue *Sciences* qui ouvre ses colonnes aux RCT (Banerjee *et alii*, 2015b). Rares sont aujourd'hui les formations académiques prétendant "tutoyer l'excellence" qui ne proposent un cursus spécialisé dans ce domaine, comme dans les grandes universités américaines (Harvard, MIT, Yale, etc.) ou ailleurs (London School of Economics, Écoles d'économie de Paris, de Toulouse ou de Marseille, etc.). Rares également les conférences internationales qui ne programment pas de session dédiée aux RCT avec un succès d'assistance réitéré. Rares enfin, les agences d'aides qui n'ont pas créé de département qui leur soit dédié ou qui n'aient engagé ou financé leurs propres RCT.

Sur le plan méthodologique et de la connaissance en économie du développement, les RCT représentent un indiscutable progrès. Néanmoins et en dépit d'un champ d'application restreint (évaluation de projets spécifiques, localement circonscrits et souvent à petite échelle), les RCT sont aujourd'hui labélisées "Gold Standard" de l'évaluation, à l'aune duquel il conviendrait de jauger toute approche alternative. Présentées par ses adeptes comme une véritable révolution copernicienne en économie du développement, on leur attribue en exclusive le qualificatif de "rigoureuses", voire de "scientifiques". Bien au-delà du champ méthodologique, l'ambition de certains défenseurs des RCT les plus médiatiques, au premier rang desquels Esther Duflo, est de fournir une liste exhaustive des bonnes et des mauvaises politiques en matière de développement. L'accumulation d'un nombre toujours plus grand d'études d'impact, sur lesquelles il conviendrait de capitaliser afin d'en tirer les enseignements généralisateurs, serait la voie fondatrice de cette ambition universalisante. Plus concrètement, la suprématie revendiquée des RCT dans le champ de l'évaluation quantitative engendre un certain nombre d'effets pervers qui seront discutés ici : disqualification et effet d'éviction des méthodes alternatives, mobilisation toujours plus grande des ressources allouées, positions de rentes. Il existe donc un véritable hiatus entre l'étroitesse du champ d'application et l'hégémonie revendiquée par les promoteurs des RCT les plus en vue. C'est ce paradoxe que nous nous proposons d'explorer dans cet article.

Dans la première partie, nous présentons brièvement les principes fondateurs des RCT et ses avantages théoriques, notamment en regard des autres méthodes d'évaluation existantes, ainsi que leur montée en puissance. La deuxième partie est consacrée à la discussion et la circonscription du champ réel d'application de ces méthodes. En identifiant leurs limites, notamment lors du passage des conditions expérimentales idéales à leur application sur le terrain, mais aussi en étudiant leurs usages (et instrumentalisations) politiques, nous tentons d'établir leur périmètre de validité dans l'ensemble plus général des questions de développement. Enfin, la troisième partie adopte une perspective d'économie politique plus large, appliqué à la méthode dans son ensemble. L'économie politique est définie ici comme les interactions entre les forces politiques (qui peuvent être le fait d'institutions, de groupes organisés ou d'individus) et les activités économiques (ici en l'occurrence les RCT, dont on montrera qu'elles sont devenues une véritable industrie) et la manière dont ces deux aspects s'influencent mutuellement. Dans cette perspective, nous nous attachons à décrypter la façon dont les différents acteurs interagissent, les jeux et enjeux de pouvoir et au bénéfice de qui ils opèrent. Cette dernière partie propose d'expliquer pourquoi ces méthodes jouissent d'une légitimité politique et académique qui va bien au-delà de leur réelle portée scientifique. Nous concluons en proposant notre propre vision en matière de méthode d'évaluation d'impact et en dressant quelques pistes de recherche dans le prolongement de cet article.

## I. La montée en puissance d'une méthodologie

Les méthodes expérimentales visent à comparer la situation issue d'un projet (d'un programme ou d'une politique) à celle qui aurait eu cours s'il n'avait pas été mis en place, afin d'en mesurer l'impact net ; c'est-à-dire, une fois purgé de tous les changements advenus par ailleurs. Toute la difficulté est de construire la situation de référence hypothétique (sans projet ; appelée *contrefactuel*), qui par définition n'est jamais observée. La solution proposée par les méthodes expérimentales consiste à sélectionner par tirage aléatoire deux échantillons au sein d'une même population susceptible de bénéficier de l'intervention. Celle-ci ne sera attribuée qu'à l'un des groupes, mais les deux feront l'objet d'enquêtes avant et après le projet. Les propriétés statistiques issues de la théorie des sondages garantissent qu'en moyenne, les différences observées entre les bénéficiaires et les non-bénéficiaires peuvent être attribuées au projet. Comme dans toute méthode probabiliste, les résultats sont connus avec une marge d'erreur (intervalle confiance) qui dépend des caractéristiques de l'échantillonnage (taille, méthode, attrition, etc.).

Les méthodes expérimentales cherchent donc à établir formellement une relation causale entre une intervention et un certain nombre de variables de résultats. D'un point de vue scientifique, et en théorie, elles peuvent être légitimement considérées comme l'alternative la plus convaincante pour identifier l'existence et quantifier l'ampleur de l'impact observé. Dans le champ des évaluations quantitatives, elles apparaissent *a priori* plus robustes que les autres méthodes existantes : en l'absence de groupe de contrôle constitué *ex ante*, les approches *avant-après* ou *sans-avec* ne permettent pas de contrôler les variations du contexte ; les méthodes quasi-expérimentales d'appariements – qui appairer bénéficiaires et non-bénéficiaires sur la base de caractéristiques observables communes – lèvent partiellement cette contrainte. Mais en l'absence

de sélection aléatoire ex ante, elles omettent les caractéristiques inobservables qui pourraient avoir influencé la sélection et qui différencieraient donc le groupe traité du groupe de contrôle (aversion pour le risque, "dynamisme entrepreneurial", insertion dans des réseaux sociaux, etc.). Toujours dans le champ quantitatif, les RCT permettent de surmonter, a priori, le défi méthodologique que constitue la démonstration du sens de la causalité, sans reposer sur des hypothèses économétriques complexes et toujours réfutables. Enfin et plus classiquement, elles se distinguent des méthodes qualitatives (études de cas, monographies, entretiens, observation participante) en proposant une mesure chiffrée de l'impact, hors de portée (et de propos) de ces dernières. La méthode expérimentale est devenue à tel point incontournable pour estimer des relations de causalité, qu'initialement opposée aux techniques économétriques, elle en est devenue le "benchmark", comme en atteste le titre du chapitre introductif ("*The Experiment Ideal*") d'un manuel très populaire d'économétrie (Angrist et Pischke, 2009).

Au-delà de ces avantages génériques (et théoriques), la transposition et la multiplication des RCT aux questions de développement sont à saluer pour d'autres raisons. Nous en citerons trois principales. En premier lieu, elles ont mis le doigt sur un point aveugle des politiques dans les PED, qu'elles soient nationales ou issues de l'aide publique au développement (APD), à savoir leur manque criant d'évaluation quantitative dans le passé. Des sommes colossales ont ainsi été dépensées sans qu'on ait une idée précise de leur efficacité, ouvrant un boulevard à des critiques radicales sur des bases plus idéologiques que scientifiques (Easterly, 2009 ; Moyo, 2009). L'acceptation du principe des évaluations et leur multiplication ne peuvent que concourir à la redevabilité démocratique aussi bien au Sud qu'au Nord (Cling *et alii*, 2003). En second lieu, les RCT ont donné un nouvel élan à la collecte de données d'enquêtes de première main par les économistes du développement. Longtemps cantonnés à la modélisation d'agrégats macroéconomiques issus des grandes bases de données internationales de qualité douteuse, tout particulièrement en Afrique (Jerven, 2013 et 2015 ; Devaradjan, 2013), les chercheurs peuvent s'appuyer sur la légitimité accordée aux RCT dans le champ du *mainstream* pour opérer un rapprochement nécessaire avec le terrain et les acteurs. Enfin, jusqu'alors, les PED étaient marginalisés par la recherche économique en raison de leur déficit de données de qualité, en particulier longitudinales. La généralisation des RCT permet de placer la recherche économique sur ces pays au niveau des meilleurs standards internationaux. Elle se présente même comme un progrès méthodologique initié au Sud et transféré vers le Nord.

Ces atouts ont conduit à une montée en puissance foudroyante des RCT dans les PED à partir du milieu des années 2000. Le Lateef Jameel Poverty Action Lab (J-PAL) a constitué l'un des promoteurs les plus influents des RCT, animant une vaste campagne de plaidoyer en leur faveur. Il a été fondé en 2003 par Abhijit Banerjee, Esther Duflo, chercheurs au *Massachusetts Institute of Technology* et Sendhil Mullainathan, chercheur à Harvard. Dédié exclusivement aux RCT, reconnu comme un label de qualité en la matière, il organise des formations et des échanges de pratiques au travers d'un réseau affiliant notamment 146 professeurs d'universités<sup>1</sup> ainsi que de nombreux chercheurs (Figure 1). Il facilite l'accès aux financements pour mener des études randomisées et

---

<sup>1</sup> Informations collectées sur le site web du J-PAL le 24/01/2017.

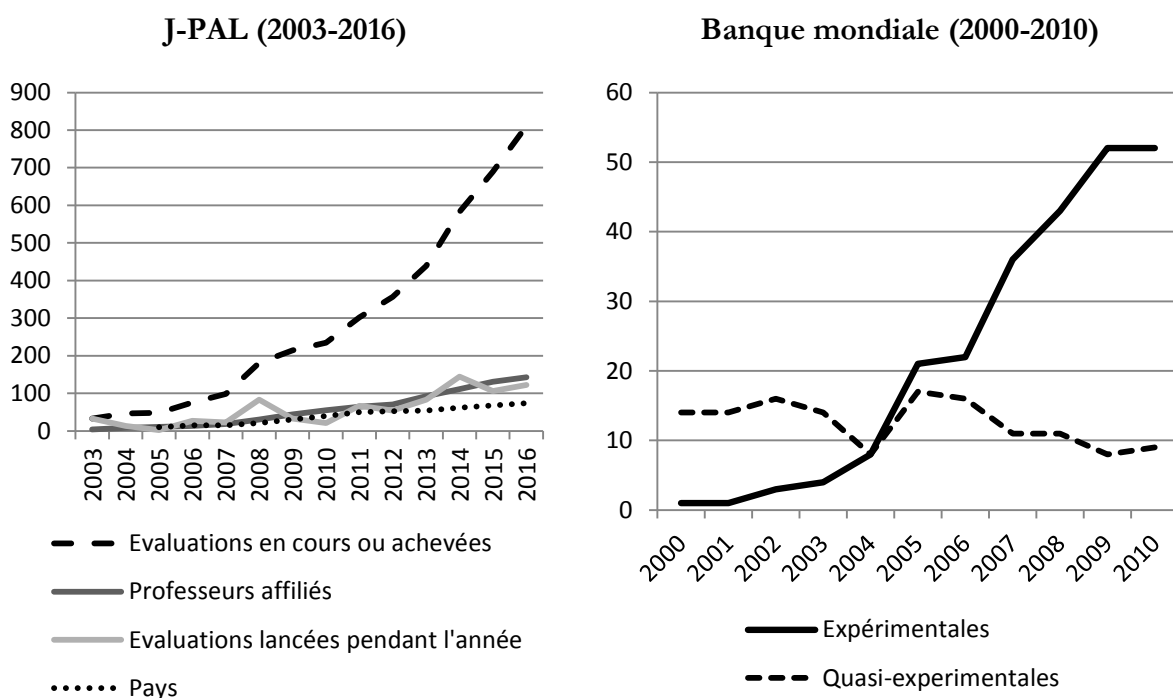
stimule la diffusion des résultats, tant dans le champ scientifique qu'à l'attention des dirigeants politiques. Ce laboratoire est étroitement lié à l'ONG IPA (*Innovations for Poverty Action*), une organisation chargée de répliquer à grande échelle les évaluations et les programmes. En janvier 2017, treize ans après sa création, le J-PAL affiche s'être engagé dans pas moins de 811 évaluations (achevées ou en cours) dans 74 pays, avec une croissance continue au fil des ans. L'Afrique arrive au premier rang des terrains d'application (avec 240 expériences), loin devant l'Asie du Sud (165, essentiellement en Inde) et l'Amérique latine (131). Les thèmes de prédilection portent sur la finance (241), suivi des secteurs sociaux (196 pour l'éducation, et 166 pour la santé), la question de la gouvernance suscitant un intérêt exponentiel (185)<sup>2</sup>. On notera le rôle central joué par Esther Duflo, qui a participé à pas moins de 49 évaluations (dont 13 en cours), elle-même largement dépassée par Dean Karlan, qui en affiche 100 (dont 42 en cours) ! Ces performances ne laissent pas d'interroger sur la proximité des chercheurs au terrain évoquée plus haut (nous y reviendrons).

Si la Banque mondiale semble avoir une politique plus nuancée en matière de méthodologie d'évaluation d'impact, en diversifiant les méthodes (néanmoins toujours quantitatives), les RCT en représentent près des deux tiers, avec 64% des 368 évaluations engagées (jusqu'en 2010). Non seulement les RCT tendent à occuper une position de plus en plus dominante, mais elles exercent un effet d'éviction sur les autres approches, comme le montre la figure 1. Au cours de la période 2000-2004, à peine 20% des évaluations étaient des RCT. Dans les six années suivantes, les proportions ont été totalement inversées (76%). Le nombre de RCT est en progression constante, tandis que les évaluations appliquant d'autres méthodes stagnent voire régressent. Ces chiffres consolidés n'ont malheureusement pas été actualisés après 2010, mais il est peu probable que la situation ait profondément changé. Outre la création du DIME (*Development Impact Evaluation Initiative* ; voir partie III) en 2005, d'autres fonds spécialement dédiés aux RCT ont été mis en place, tels que le *Strategic impact Evaluation Fund* (SIEF) en 2007, le *Global Agriculture and Food Security Program* (GAFSP dont 30% des interventions doivent faire l'objet d'expérimentation) en 2009 ou l'*Impact Evaluation to Development Impact* (I2I) en 2014. En 2016, Esther Duflo indiquait qu'environ 400 projets en cours financés par la Banque mondiale font l'objet de RCT (Duflo, 2016).

---

<sup>2</sup> Dans la dernière version de la base d'évaluation accessible sur le site du J-PAL que nous avons utilisée, une évaluation peut se rapporter à plusieurs thèmes.

Figure 1: Évolution des évaluations d'impact lancées par J-PAL et la Banque mondiale



Source : Jatteau (2016) pour 2003-2015, et site J-PAL pour 2016 (actualisé le 06/01/2017) ; IEG, 2012.

L'évaluation d'impact a ainsi connu un succès fulgurant au cours des dernières années, au point de devenir une véritable industrie, les RCT s'adjugeant la part du lion : suite à l'actualisation de son *Impact Evaluation Repository*, le 3IE<sup>3</sup> recensait en août 2016, plus de 4 260 évaluations d'impact achevées ou en cours dans le domaine du développement, dont 2 645 RCT (Miranda *et alii*, 2016 ; Figure 2). La question de savoir si l'âge d'or du « gold standard » est derrière nous reste ouverte (nous y reviendrons en conclusion). La Figure 2 semble indiquer que le nombre de RCT en chantier marque le pas. Mais les chiffres se maintiennent à un haut niveau, tandis que d'autres indicateurs ne font pas état d'une telle stagnation (voir Figure 1 par exemple), ce qui soulève la question de savoir si cette dernière n'est pas l'effet d'un recensement encore incomplet pour les années les plus récentes.

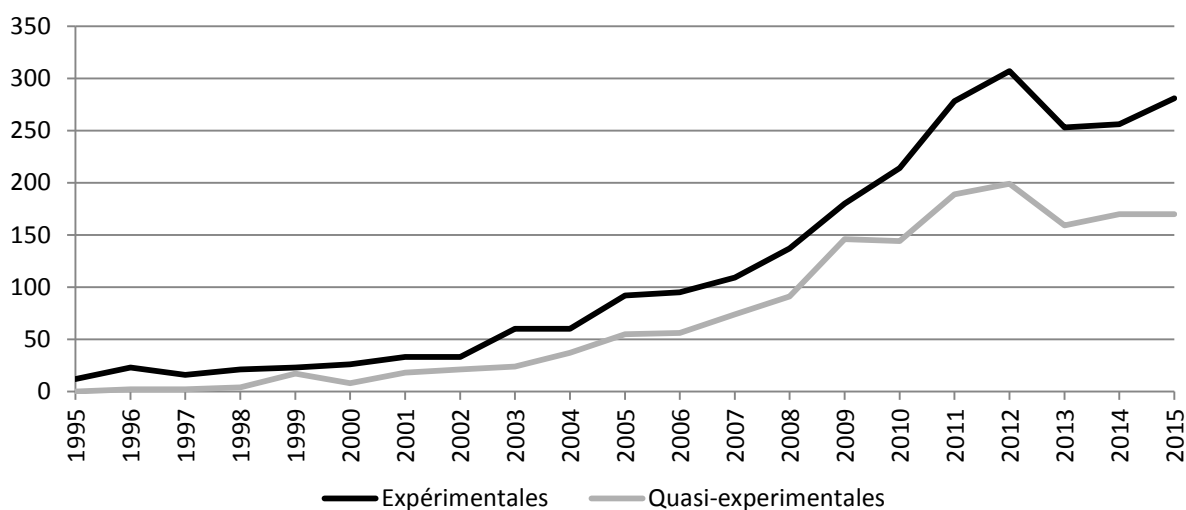
L'industrie des RCT, qui s'est envolée au cours des dix dernières années, reste à ce jour florissante et domine largement le champ de l'évaluation d'impact des politiques de développement. Elles ont essaimé partout dans le monde, et ont mobilisé des ressources considérables. S'il n'est pas possible de quantifier les flux financiers en jeu faute de données, ils se comptent en centaines de millions de dollars). A titre d'exemple, si le J-PAL ne publie pas de rapport financier, les revenus annuels d'IPA sont passés de 252 000 dollars en 2003 à plus de

<sup>3</sup> 3ie (*International Initiative for Impact Evaluation*) est une ONG internationale créée en 2008 et soutenue par de grandes fondations et des bailleurs de fonds, spécialisée dans la promotion de programmes et de politiques de développement « fondés sur les preuves » (*evidence based*) dans les pays en développement. En septembre 2016, elle avait octroyé 108,8 millions de dollars de subvention à des études, principalement pour financer 216 évaluations d'impact (source : <http://www.3ieimpact.org/en/about/performance-metrics/>, actualisé en septembre 2016 et consultée le 02/02/2017).



39 millions en 2015<sup>4</sup>. Les RCT ont aussi donné lieu à de nombreux manuels de bonnes pratiques et des dizaines d'articles académiques, en particulier dans les revues les plus prestigieuses : en 2015, les RCT constituaient 31% des articles en économie du développement publiés dans les revues dites du « top five »<sup>5</sup> et plus de 40% dans les revues d'intérêt général classées ensuite<sup>6</sup> (McKenzie, 2016). Il oriente aussi les vocations : 65% des économistes associés au BREAD qui ont eu leur doctorat depuis 2006 ont participé à au moins une RCT (Duflo, 2016). Face à ce phénomène saisissant (on parle d'un "mouvement pro évaluation d'impact"), il convient de s'interroger sur cet engouement.

**Figure 2: Evolution du nombre d'évaluations d'impact dans le champ du développement par type de méthode (1995-2015)**



Source : 3ie Impact Evaluation Repository, mis à jour en Août 2016.

Une telle systématisation des RCT est-elle scientifiquement légitime et politiquement souhaitable ? Deux questions font problème : d'une part la revendication de la supériorité intrinsèque des RCT sur toute autre méthode (Duflo *et alii*, 2007) ; d'autre part l'idée que l'accumulation tous azimuts des RCT permettra, par effet de masse, de répondre à toutes les questions de développement, sur "ce qui marche et ce qui ne marche pas" sur des bases incontestables (Duflo et Kremer, 2005) ; des revendications qui n'ont pas été abandonnées depuis par les promoteurs du mouvement. L'audition de Karlan (2015) par le Congrès américain en est la plus parfaite illustration.

## II. Critiques de la méthode : de la théorie à la mise en œuvre

Les tenants des RCT en économie du développement les ont importées du champ médical en faisant abstraction des discussions critiques, des conditions de mises en œuvre et remises en cause

<sup>4</sup> Rapports financiers bisannuels disponibles à la page [www.poverty-action.org/about/annual-reports-finances](http://www.poverty-action.org/about/annual-reports-finances), consultée le 24/01/2017.

<sup>5</sup> Le "Top 5" selon IDEAS-RePEC : *The Quarterly Journal of Economics*, *Journal of Political Economy*, *American Economic Review*, *Econometrica*, et *Journal of Economic Literature* (liste au 02/02/2017).

<sup>6</sup> *American Economic Journal: Applied Economics*, *Economic Journal* et *Review of Economics and Statistics*.

dont elles ont fait l'objet en santé publique (Labrousse, 2010 ; Eble *et alii*, 2014). Ils ont également fait l'impasse sur les controverses qui avaient marqué plusieurs décennies de débats en économie du développement (Picciotto, 2012). Nous rappelons ici de manière synthétique les principales critiques adressées aux RCT, tant du point de vue de la validité interne qu'externe, avant de nous pencher sur les implications de la confrontation au terrain. Nous montrerons, à partir d'exemples concrets, que l'économie politique des RCT a en pratique des conséquences de premier ordre qui remettent en question les propriétés théoriques revendiquées de la méthode.

## 2.1. Discussion de la validité (interne et externe) des RCT

La validité interne des RCT (c'est-à-dire la fiabilité des résultats obtenus), est censée être le principal point fort de la méthode. Pourtant, les RCT sont loin d'être exemptes de limites. Si ces dernières sont surtout de nature empirique, dans la manière dont les RCT sont conduites sur le terrain, certaines sont d'ordre plus théorique. Sur ce plan, la principale critique porte sur la tension entre biais (à minimiser) et précision (à maximiser), que les RCT ne permettent pas d'arbitrer de manière optimale, suivant les principes de base de la théorie de la décision (Deaton et Cartwright, 2016)<sup>7</sup>. Les analyses randomisées se focalisent sur les résultats moyens, pour l'ensemble de la population considérée. Si par principe, les RCT fournissent des estimateurs sans biais de l'effet moyen, elles ne garantissent pas qu'ils soient de variance minimale ; pas plus qu'elles ne permettent de calculer l'effet médian, ou par quantile. C'est particulièrement problématique lorsque les effets du traitement sont hétérogènes. Or, en médecine, et plus encore en matière de développement, les patients, ou usagers, sont susceptibles de réagir très diversement à un traitement. En l'absence d'information sur la variance, on ne dispose d'aucune méthode robuste pour quantifier les marges d'erreurs des RCT et les tests utilisés sont en fait le plus souvent inadéquats. La précision pourrait être améliorée, par des méthodes de stratification par exemple, mais cela suppose une connaissance préalable minimale sur le contexte à laquelle nombre de randomisateurs se refusent. D'une certaine façon le discours sur la supériorité de la randomisation se prend à son propre piège : pour améliorer les estimations il faudrait pouvoir être capable de faire des hypothèses raisonnées (*assumptions*), alors que le mouvement en faveur des RCT a justement fondé la supériorité de cette méthode sur son absence de *priors*, ne mobilisant rien de plus que la loi des grands nombre. Finalement, les RCT ne résolvent en rien les problèmes classiques de l'inférence statistique (notamment en présence de distribution asymétrique, d'*outliers*, etc.). La pratique courante qui consiste à se départir de la comparaison simple de moyenne entre groupe traité et contrôle par l'estimation de régressions économétriques n'est pas une solution, au contraire, comme le montre Young (2016). Après avoir appliqué un test approprié à 2003 régressions présentées dans 53 papiers issus de RCT publiés dans les meilleures revues américaines, il trouve que 30% à 40% des coefficients estimés significatifs ne le sont en fait pas. Toutes ces difficultés cumulées conduisent Deaton et Cartwright (2016) à conclure de manière tranchée : « *it is almost never the case that an RCT can be judged superior to a well-conducted observational study simply by virtue of being an RCT* ».

---

<sup>7</sup> Voir aussi, entre autres, Heckman (1991), Deaton (2010), Ravallion (2009), Barrett et Carter (2010), Rodrik (2008).

Sur le plan empirique, les RCT n'échappent pas au « bricolage » habituel des protocoles de recherche, tout particulièrement en sciences sociales. De nombreuses contraintes de *mise en œuvre* sapent les fondements mêmes de l'échantillonnage aléatoire, et donc les soubassements de la prétendue supériorité scientifique des RCT. On peut citer ici : l'absence de véritable tirage aléatoire, pour des raisons éthiques ou pratiques (ne pas perturber la mise en œuvre d'un projet) (Scriven, 2008 ; Labrousse, 2010 ; Deaton, 2010) ; des variables de traitements souvent trop mal spécifiées pour pouvoir servir d'estimateurs structurels (Heckman et Vytlačil, 2005) ; l'absence de double aveugle, qui introduit des incitations diverses à la participation – ou à la non-participation (Heckman *et alii*, 1998 ; Rosholm et Skipper, 2009) ; lorsque celle-ci est trop faible, des mesures diverses de la part des chercheurs visant à « forcer » la participation, le programme évalué s'éloignant alors sensiblement du programme « normal » (Bernard *et alii*, 2012 ; Quentin et Guérin, 2013 ; Morvant-Roux *et alii*, 2014) ; des effets de contamination mais aussi d'attrition, lesquels, malgré des efforts multiples<sup>8</sup>, peinent à être résolus (Eble *et alii*, 2014). Soulignées de longue date en médecine, ces difficultés ont depuis été mises en évidence dans de nombreux exemples concrets, sans pour autant que les *randomistas* fassent davantage preuve de réflexivité dans leurs protocoles de recherche (Shaffer, 2011). Ainsi, les propriétés méthodologiques des RCT conduite dans le champ du développement sont systématiquement inférieures à celles conduites en médecine (au Nord comme au Sud) ou dans les pays développés pour les politiques sociales (Eble *et alii*, 2014).

La question de la validité externe est de loin celle qui a été discutée dans la littérature. La focalisation sur un impact « moyen », outre les problèmes statistiques mentionnés plus haut, ne nous dit rien sur l'hétérogénéité des impacts et leur distribution (Ravallion, 2009 ; DFID, 2012). Se limiter à un impact de court terme (pour des raisons de coût et d'attrition) amène souvent à étudier des indicateurs intermédiaires, qui peuvent être très différents des résultats finaux (Eble *et alii* ; 2014), voire contradictoires puisque nombre de projets ne suivent pas une trajectoire linéaire (Labrousse 2010 ; Woolcock 2009). Les effets d'entraînement et d'équilibre général, pourtant légions, sont ignorés (Acemoglu, 2010 ; Ravallion, 2009 ; Deaton and Cartwright, 2016)<sup>9</sup>. Il en va de même de la dimension politique de réplication des programmes, qui en est pourtant une composante essentielle (Bold *et alii*, 2013 ; Pritchett et Sandefur, 2013 ; Acemoglu, 2010). Enfin les *raisons* de l'impact sont ignorées : les RCT permettent éventuellement de mesurer certains impacts ou de tester plusieurs modalités d'une intervention, mais elles ne permettent pas d'en analyser les *mécanismes* ni les *processus* sous-jacents. Dépasser cette limite de la causalité probabiliste supposerait un « modèle causal » (Cartwright, 2010), une théorie du changement cohérente (Woolcock, 2013), une approche structurelle (Acemoglu, 2010) ainsi qu'une contextualisation

---

<sup>8</sup> Pour un exemple de travaux proposant des méthodes pour essayer de corriger les effets d'attrition (ainsi que d'autres biais), voir Duflo *et alii* (2007). Mais ces recommandations ne sont en général pas suivies dans la pratique (Eble *et alii*, 2014).

<sup>9</sup> Pour l'anecdote, on relèvera le papier de Cahuc et Le Barbanchon (2009), très critique vis-à-vis des RCT pour ignorer les effets d'équilibre général dans l'évaluation des politiques publiques. Ce même auteur sera à l'origine d'une immense polémique en France, en co-signant fin 2016 un ouvrage virulent contre tous les économistes, qualifiés de « négationnistes » dont il faudrait se débarrasser (selon le titre de l'ouvrage), qui ne considèrent pas leur discipline comme une science expérimentale, et dont les RCT seraient le gold standard (Cahuc et Zylberberg, 2016).

(Ravallion, 2009 ; Pritchett et Sandefur, 2015). Ce n'est pas le moindre des paradoxes que les RCT qui sont présentées par ses promoteurs comme les seules à pouvoir identifier un véritable impact causal, n'aient finalement absolument rien à dire sur les processus à l'œuvre, c'est-à-dire les causes de l'impact.

Les RCT, quel que soit leur champ d'application, sacrifient la validité externe au détriment de la validité interne (Cartwright, 2010). En matière de décisions politiques, cet arbitrage est une erreur, suggèrent Pritchett et Sandefur (2015). A partir d'exemples pris dans le champ de l'économie de l'éducation (l'impact de la taille des classes et le rendement de l'éducation), puis du microcrédit, les deux auteurs suggèrent qu'il est beaucoup plus pertinent de prendre des décisions politiques dans un contexte donné en s'inspirant d'études non randomisées mais menées dans le même contexte que d'études randomisées menées ailleurs. Montant en généralité, ils s'emploient à démontrer formellement que la revendication de validité externe de l'impact estimé des RCT est nécessairement erronée et que les recommandations de politique qu'on en tire sont mal fondées. Empruntant une autre approche, celle de l'économie politique, O'Laughlin (2015) aboutit à un résultat convergent dans le domaine épidémiologique et plus précisément celui du HIV, à partir du cas de l'Afrique Australe.

L'ensemble de ces éléments rend difficile toute exploitation pour améliorer ou concevoir des politiques à partir de RCT, ce qui en limite fortement le champ d'application. Une fois que l'on a admis l'ensemble de ces critiques, tant en termes de validité interne qu'externe, que reste-t-il aux RCT ?

Deaton et Cartwright (2016) suggèrent que les RCT restent néanmoins valables dans deux domaines : 1) tester une théorie, 2) évaluer ponctuellement et dans un contexte donné un projet ou une politique particulière, mais à condition que les problèmes potentiels de validité interne soient résolus. Or ceci nous ramène à un spectre très étroit (Picciotto, 2012), que Bernard *et alii* (2012) qualifient de programmes « tunnels ». Ces derniers se caractérisent par des impacts de court terme, des inputs et outputs clairement identifiés, facilement mesurables, des liens de causalité unidirectionnels (A cause B), linéaires et enfin non soumis à des risques de faible participation de la part des populations visées. Ils rejoignent les suggestions de Woolkcock (2013) : les projets qui se prêtent à la randomisation doivent avoir une « low causal density », être peu contingents aux capacités de mise en œuvre (*implementation capability*) et comporter des résultats prédictibles.

L'ensemble de ces conditions exclut un grand nombre de politiques de développement, qui mettent en jeu des combinaisons de mécanismes socioéconomiques et des boucles de rétroaction (effets d'émulation, d'apprentissage des bénéficiaires, d'amélioration de la qualité des programmes, effets d'équilibre général, etc.). Dans les termes de référence d'une étude commanditée sur le sujet, certains responsables du DFID estimaient ainsi le champ d'application des RCT à moins de 5% des interventions de développement (DFID, 2012). S'il convient de ne pas prendre ce chiffre au pied de la lettre, il ne fait aucun doute que les méthodes expérimentales ne sont pas adaptées pour évaluer l'impact de la grande majorité des politiques de

développement. Dans leur papier plus formalisé, Sandefur et Pritchett (2013) aboutissent à des conclusions similaires<sup>10</sup>.

Circonscrire le champ des évaluations d'impact aux interventions susceptibles de respecter les canons de la randomisation, non seulement écarte un grand nombre de projets, mais aussi nombre de dimensions structurelles du développement, tant économiques que politiques, comme la régulation des grandes compagnies, la fiscalité, les échanges internationaux, pour n'en citer que quelques unes. Face à des causalités souvent complexes et systémiques, et plutôt que de s'interroger sur « ce qui marche », ne faut-il pas s'interroger sur les moyens de promouvoir des projets flexibles, adaptatifs et créatifs, comme le suggèrent Pritchett *et alii* (2013) ? Cela implique d'inclure la collecte de données utiles au sein des projets, puis de mettre en place des systèmes de suivi permettant un apprentissage itératif, permettant progressivement d'ajuster le projet à un environnement à la fois complexe et volatil.

Reconnaissant que les résultats des RCT restent très liés aux contextes chaque fois spécifiques où elles sont appliquées (période, lieu, modalités d'intervention du projet), certains de leurs promoteurs les plus en vue arguent qu'il convient de les considérer comme des « biens publics mondiaux » et de créer une instance internationale chargée de les multiplier (Svedoff *et alii*, 2006 ; Glennerster, 2012). Celle-ci constituerait ainsi une base de données universelle et jouerait le rôle de « chambre de compensation » apportant des réponses sur tout ce qui marche ou ne marche pas en matière de développement (Duflo et Kremer, 2005 ; Banerjee et Hee, 2008) . Ce projet généralisateur s'illustre sur le site de J-PAL, qui met en avant 11 projets évalués ayant fait l'objet d'une massification – formation de la police pour la thématique « économie politique et gouvernance », vermifuges et soutien scolaire pour l'éducation, distribution gratuite de moustiquaires pour la santé – avec 202 millions de personnes bénéficiaires.

Les RCT se focalisent sur de petits dispositifs, relativement simples et facilement actionnables qui ne sauraient, mises bout à bout, retranscrire l'intégralité des enjeux du développement ou fonder une politique sociale. Les limites évoquées plus haut en matière de validité externe rendent illusoire la prétention des randomisateurs à proposer un panier de politiques globales sur la base de RCT nécessairement localisées. C'est d'autant plus vrai, qu'il n'y a pas de loi en sciences sociales comme il en existe en sciences dites "dures" (physiques ou naturelles). Il n'existe donc pas de paramètres universels à estimer équivalents aux constantes gravitationnelles, d'Euler, etc. La question d'économie politique du passage à échelle (par exemple au niveau national) de politiques évaluées dans des conditions expérimentales, et donc la nécessité de mobiliser les institutions publiques structurellement faibles, nous paraît particulièrement difficile à résoudre (voir ci-dessous). Ainsi, François Bourguignon, chercheur renommé ayant largement contribué à promouvoir les RCT, considère cette proposition absurde et scientifiquement infondée<sup>11</sup>. Dans

---

<sup>10</sup> « *The scope of application of "planning with rigorous evidence" approach to development is vanishing small* » (Sandefur et Pritchett, 2013, p.1)

<sup>11</sup> François Bourguignon a été directeur de l'École d'économie de Paris entre 2007 et 2013. Précédemment, il a été économiste en chef et premier vice-président de la Banque mondiale à Washington entre 2003 et 2007, période au cours de laquelle il a contribué à la création du DIME. Dans son discours

ces conditions, la poursuite de ce projet pharaonique est au mieux une fuite en avant, mais plus probablement le fruit d'intérêts qu'il convient d'identifier. Nous commencerons par ceux qui sont en jeu lors des évaluations sur le terrain.

## 2.2. Quelle portée (en pratique) : l'économie politique de quatre RCT *in the field*

La sociologie des sciences, autour des travaux de Bruno Latour, montre à quel point la production puis l'utilisation de résultats scientifiques, quels qu'ils soient, ne peuvent être dissociés des dynamiques sociopolitiques qui les entourent. Les résultats, y compris des résultats chiffrés dont on pourrait imaginer qu'ils soient irréfutables et utilisables tels quels, n'échappent pas à cette fabrique politique. Leur mise en mots (le passage de l'économétrie au texte) suppose un premier processus de traduction. Leur diffusion puis leur réception en suppose un second : ils sont alors réappropriés, transformés, parfois contournés ou subvertis, selon des logiques diverses et difficilement prévisibles puisqu'elles dépendent de contextes historiques, sociaux et politiques singuliers. Comprendre l'impact des RCT ne peut faire l'économie de ce type d'analyse. En d'autres termes, ce sont les RCT elles-mêmes qui méritent d'être passées au crible d'une analyse d'économie politique. La partie suivante est consacrée à l'économie politique de la méthode dans sa globalité. Peu d'éléments empiriques sont disponibles sur la manière dont sont effectivement conduites les expérimentations à cette échelle : ils supposeraient de pouvoir analyser les processus de mise en œuvre de ces évaluations, lesquels sont souvent mal documentés et dont les observateurs extérieurs ou critiques sont souvent écartés<sup>12</sup>. Nous avons toutefois pu rassembler des éléments très précis sur quatre RCT emblématiques, effectués par les chercheurs les plus renommés du domaine, publiés dans les meilleures revues et abondamment cités. Ces exemples nous permettront d'illustrer l'économie politique de ces évaluations, c'est-à-dire la manière dont celles-ci, indépendamment de leur rigueur (interne ou externe), se prêtent à des influences sociopolitiques multiples, tant dans leur formulation et leur mise en place que dans la diffusion de leurs résultats.

Commençons par le cas emblématique de l'évaluation par l'IFPRI du programme *Progresa*, renommé ultérieurement *Oportunidades* puis *Prospera*. Mis en place en 1997 par le gouvernement mexicain, il visait à des offrir des transferts monétaires à des ménages pauvres si ces derniers suivaient un certain nombre de prescriptions visant à améliorer l'éducation et la santé de leurs enfants. Ce programme est à l'origine de la généralisation des politiques de transferts conditionnels (*Conditional Cash Transfers* ; CCT) dans les PED à partir de la fin des années 90. Cet exemple montre comment une évaluation, en dépit d'une validité interne médiocre mais passée

---

de clôture de la conférence de AFD-EUDN à Paris le 26 mars 2012, il déclarait (extraits) : « *There has been this fashion during the last couple of years on the RCTs. We even heard colleagues, good colleagues, saying that in the field of development, and in the field of development aid, the only fruitful approach from now on was to do random control trials in all possible fields of interventions. And at the end, we'll have a huge map, a huge catalogue saying "This works, this doesn't work". This is crazy! This will never work and, because of that, we absolutely need the other approaches to evaluating policies and programs. The "pure, scientific evidence" on all what is concerned with development is simply completely impossible. We have to live with this imperfect knowledge* » (souligné par nous).

<sup>12</sup> Voir Jatteau 2016, pp. 56-66.

sous silence, peut avoir une portée considérable en matière de politiques publiques (et de légitimation de la méthode).

Par un travail minutieux d'exhumation de la documentation technique (mais non publiée) de l'évaluation, Faulkner (2014) montre qu'aussi bien le protocole initial (choix de l'échantillon) que sa mise en œuvre (phénomène d'attrition et de contamination) s'éloignent substantiellement du cadre théorique des RCT. Par exemple sur le premier aspect, les échantillons initiaux des groupes de traitement et de contrôle ont été choisis dans deux univers différents, et non tirés aléatoirement comme l'impose le principe même de la randomisation. Ce qui n'empêche pas les publications ultérieures de présenter le protocole d'échantillonnage comme vraiment expérimental. Seul l'oubli progressif de ces défaillances a permis de "vendre" l'évaluation de *Progresa* comme l'exemple fondateur du bienfondé des RCT pour estimer l'impact causal des programmes sociaux dans les PED.

Comme le souligne l'auteur, mettre l'accent (ou simplement mentionner) les points faibles de l'évaluation aurait probablement été contreproductif compte tenu de l'enjeu: à savoir la promotion internationale des RCT comme l'instrument à la fois original et le plus pertinent pour évaluer l'impact des programmes de développement ; et aussi dans le contexte mexicain, le maintien du programme suite à l'élection présidentielle en vue à la fin des années 1990, et l'alternance qui se profilait. Tous les acteurs impliqués dans le programme (chercheurs, promoteurs et décideurs) avaient intérêt à ce que l'étude soit irréprochable sur le plan de la méthode et convaincante sur le plan des résultats (Faulkner 2014 : 239).

Certains auteurs vont même jusqu'à considérer que non seulement la modalité expérimentale du protocole d'évaluation mais également le fait que les effets positifs mesurés découleraient du caractère conditionnel du programme ont été abusivement revendiqués *a posteriori*. Ceci aurait répondu à des calculs politiques, afin de sécuriser le programme après l'alternance démocratique, l'évaluation montrant même de surcroît que le programme était *cost ineffective* pour accroître le taux de scolarisation (Pritchett, 2012; Shah *et alii*, 2015).

Le deuxième exemple est issu de nos propres observations<sup>13</sup>. Financée par un bailleur français (l'AFD), réalisée par une équipe nord-américaine (Center of Evaluation for Global Action – UC Berkeley) en partenariat avec un bureau de consultant local (Domrei), cette évaluation porte sur le programme de micro-assurance SKY au Cambodge. Celui-ci est opéré par une ONG locale avec le soutien technique d'une ONG française (GRET). Une « évaluation de l'évaluation » a été menée en s'appuyant sur une reconstitution *a posteriori* des échanges emails, rapports d'étape, comptes-rendus de réunions ainsi que d'entretiens auprès des principaux protagonistes de l'étude. Comme dans n'importe quel projet de développement (depuis l'élaboration des premiers termes de références jusqu'à la diffusion des résultats l'étude aura duré plus de 5 ans et coûté plus d'un

---

<sup>13</sup> L'une d'entre nous a réalisé l'étude, détaillée dans Quentin et Guérin (2013) et les deux autres auteurs du présent article étaient membres du comité de pilotage de l'étude.

million de dollars), la complexité des jeux d'acteurs, leurs intérêts divers et parfois incompatibles ont largement modelé la nature même de l'étude.

Le bailleur, principal financeur du projet et commanditaire de l'étude, était d'abord motivé par l'idée de tester la faisabilité des études randomisées, plus que l'évaluation d'impact proprement dit, (même s'il n'y était bien sûr pas indifférent, notamment au moment de la diffusion des résultats). Les deux ONG, en quête de financement, n'avaient pas eu le choix, et souhaitaient surtout éviter que l'étude ne perturbe leurs activités quotidiennes (même si là aussi, elles seront très attentives au contenu des résultats au moment de la diffusion). Enfin l'équipe de recherche entendait, via cette étude, contribuer en priorité aux débats académiques sur la sélection adverse en matière d'assurance, sujet encore vierge concernant les pays du Sud et donc très prometteur en matière de publication. Cette diversité de priorités, inévitable dans un projet de cette envergure, aura pesé tout au long de l'étude. Source de tensions et de conflits répétés, elle aboutit à des compromis multiples, tant au niveau des objectifs de la recherche, de l'échantillonnage, de l'élaboration des questionnaires, que de l'interprétation des résultats et leur diffusion. Ces compromis ont surtout été en faveur (mais pas uniquement) de l'équipe académique américaine, probablement du fait de son prestige (dans le contexte français et cambodgien) mais aussi de l'expertise que supposait une contribution au débat sur la méthode. Deux leçons peuvent être tirées.

D'une part, ici encore, le décryptage du protocole d'enquête questionne sérieusement la validité interne de l'évaluation, passée sous silence dans la publication finale (Levine *et alii*, 2016). Suite aux exigences éthiques et pratiques des deux ONG (qui refusent de perturber leurs règles d'octroi de la microassurance, et qui, ici, ont réussi à imposer leur point de vue), le tirage aléatoire se limite à un tirage parmi les volontaires à des réunions de village. Un décalage persistant est observé entre le nom des personnes tirées au sort et celui des personnes enquêtées, ce qui laisse penser que des tricheries ont pu avoir lieu. Du fait d'un *take up* très bas, des mesures spécifiques de sensibilisation, d'incitation auprès du personnel de terrain et de rabais sur les prix ont été mises en place à l'instigation des chercheurs, le programme évalué s'éloignant alors sensiblement du programme « normal ».

D'autre part, et c'est là le point le plus instructif, la présentation et la diffusion des résultats ne peut se comprendre qu'à la lumière du jeu d'acteurs. La faible participation et les taux élevés d'abandon sont l'une des conclusions saillantes, qui confirme les observations des deux ONG depuis le début des années 2000. Concernant l'impact proprement dit, le résultat majeur concerne la diminution significative des dépenses de santé et de l'endettement des ménages, notamment ceux qui ont subi des chocs importants de santé. Aucun effet significatif n'est observé en matière d'impact sur la santé des assurés, sans pour autant que l'on puisse conclure sur ce point : le pas de temps limité de l'étude et la très faible occurrence de gros chocs de santé empêchent de déceler des effets statistiquement significatifs. La question de l'impact sur la qualité des soins, objectif pourtant prioritaire des ONG<sup>14</sup> et du bailleur, avait rapidement été écarté de l'étude car jugé

---

<sup>14</sup> Même question ici : pourquoi une seule ONG ?



« non mesurable ». De manière intéressante, les rapports d'évaluation et la présentation publique qui en sera faite à plusieurs occasions, et notamment à Phnom Penh en présence de décideurs publics, soulignent certains aspects favorables et omettent les résultats plutôt adverses, illustration des compromis qui ont fini par s'instituer entre les différents acteurs. Ils évitent surtout de poser la question centrale qui constitue pourtant l'enjeu clé pour les ONG : privilégier des assurances volontaires et payantes plutôt que des régimes obligatoires ou de gratuité ? Dans l'article académique final<sup>15</sup>, cette question de l'adhésion et de l'élasticité de l'adhésion au prix est mentionnée, mais est d'une importance très secondaire par rapport à la question des impacts.

Le troisième exemple concerne les difficultés de passage à l'échelle, d'une intervention ciblée au niveau d'une localité à une région, voire à un pays, rendant problématiques les tentatives de montée en généralité à partir d'un programme circonscrit dans l'espace. La question du "*scaling up*" n'est pas seulement un problème technique (externalités, contamination, saturation, effet d'équilibre général, etc.), c'est également un problème d'économie politique. Un programme de contractualisation d'enseignants vacataires (plutôt que fonctionnaires) au Kenya est emblématique à cet effet. Une RCT de ce programme expérimenté à petite échelle par une ONG avait montré un impact positif sur le niveau d'éducation des élèves (Duflo *et alii*, 2012). Ces résultats positifs semblaient d'autant plus robustes qu'ils venaient conforter ceux obtenus en Inde sur un programme similaire (Muralidharan et Sundararaman, 2011).

Néanmoins, Bold *et alii* (2013) ont pu montrer, que ce programme n'avait plus aucun effet une fois généralisé et mis en œuvre par l'Etat. Basée sur la mise en place d'une RCT très originale (la première du genre à tester des effets organisationnels et d'économie politique) qui tire avantage de l'extension d'un programme au niveau national, l'étude de Bold et ses co-auteurs conclut que l'absence d'effet une fois passé à l'échelle s'explique par le changement d'opérateur du projet : des ONG soigneusement sélectionnées et très motivées, d'un côté ; des fonctionnaires gouvernementaux syndiqués de l'autre. Ce biais pourrait même être systématique s'il existait une corrélation entre les lieux, personnes et organisations qui acceptent de mettre en œuvre des RCT et les impacts estimés (Pritchett et Sandefur, 2013). L'argument théorique d'Acemoglu (2010) – qui souligne les effets d'économie politique des programmes à grande échelle de la part des groupes dont les sources de rente seraient menacées par les réformes – nous paraît particulièrement déterminant. Vivalt (2016) montre qu'il ne s'agit pas seulement d'un argument théorique ou du seul cas présenté ici : un des principaux résultats de sa méta-analyse est que les programmes gouvernementaux ont des effets significativement plus faibles que ceux mis en œuvre par les ONG ou les centres de recherche. Cet exemple pose une double limite à la généralisation de résultats de RCT portant sur des politiques conduites localement et qu'on voudrait passer à l'échelle : d'une part, le type d'opérateurs qui conduisent les politiques est déterminant (ici des institutions publiques versus des ONG) ; d'autre part, la multiplication de RCT conduites sur des terrains différents (ici l'Inde et le Kenya) ne rend en rien les résultats plus

---

<sup>15</sup> L'article est co-signé avec le consultant local, pourtant très critique à l'égard du bien-fondé de la méthode et de sa mise en œuvre, ce qui est probablement une autre illustration des compromis qui ont finis par être établis.

robustes, car le facteur déterminant de la réussite du programme est justement le type d'opérateurs qui l'a mis en œuvre.

On notera que même dans l'article pourtant publié trois ans après la controverse suscitée par cette expérience kenyane, Duflo et ses co-auteurs (2015) n'ont amendé leurs conclusions qu'à la marge. Ils ne mentionnent qu'incidemment et en conclusion le travail de Bold *et alii* (qui lui n'a toujours pas été publié en dépit de son intérêt majeur), alors que ce dernier questionne à la fois la validité externe des RCT, mais aussi implicitement la validité interne : les acteurs du traitement – les enseignants – connaissant leur statut et manipulent l'expérience).

Notre quatrième est l'évaluation du programme de vermifugation (*deworming*) d'enfants kenyans conduite à la fin des années 1990, et dont les retombées sont encore massives aujourd'hui. Dans un article célèbre publié par la prestigieuse revue *Econometrica*, Miguel et Kremer (2004), deux économistes de Berkeley et Harvard, concluent que la vermifugation a un impact significatif sur la santé et la participation scolaire, mais pas sur les résultats aux tests de connaissance. Si les auteurs identifient un impact direct sur les enfants traités, ils insistent surtout sur l'impact indirect : un effet d'externalité positive réduisant la contamination des enfants non traités au contact des premiers. C'est ce mécanisme qui est souligné dans le titre de l'article et qui a probablement été à l'origine de la publication dans *Econometrica*. De l'analyse coût-bénéfice, ils concluent que c'est la mesure la plus efficace pour améliorer la participation scolaire.

Cette étude est l'un des étendards du mouvement RCT. Elle est directement à l'origine de la campagne « Deworm the World », qui a récolté 2,5 millions de dollars en 2013 et 15,8 en 2014 (dernier compte disponible). Cette RCT fondatrice est revendiquée par J-PAL comme celle qui a fait l'objet du plus grand nombre de passages à l'échelle (5 sur 11), et qui aurait touché le plus de monde : 95 millions des 202 millions affichés (Duflo, 2016).

Longtemps considérée comme la *success story* par excellence du mouvement RCT, cette étude a récemment fait l'objet d'une véritable polémique au retentissement médiatique international (Boseley, 2015). En effet, dans le cadre d'un programme de réplique financé par le 3IE, des épidémiologistes du prestigieux laboratoire LSHTM ont contesté les résultats de l'étude initiale. En retraitant les micro-données, ils montrent dans un premier article qu'un certain nombre de résultats ne tiennent pas (Aiken *et alii*, 2015). Ils relèvent ainsi de nombreuses erreurs dans le traitement des données, notamment un problème massif de données manquantes. Après les avoir prises en compte, l'effet global (direct et indirect) de réduction de l'absentéisme scolaire est deux fois moindre qu'annoncé. Et celui-ci n'est plus significatif : si l'effet direct sur les enfants traités garde un impact, il n'y a pas d'externalité positive sur les écoles non traitées voisines, le résultat phare de l'étude initiale.

Dans un second papier, Davey et ses co-auteurs (2015), retraitent les données en appliquant des méthodes statistiques d'estimation de l'impact que les épidémiologistes considèrent comme les plus appropriées pour traiter le cas précis du protocole de randomisation effectivement mis en œuvre par le projet. Ils concluent qu'il y a bien un impact du projet sur l'absentéisme scolaire,

mais que celui-ci est faible et probablement biaisé. De plus, il n'est pas possible de l'attribuer uniquement à la vermifugation, l'autre composante du projet (éducation aux pratiques de santé) pouvant être seule à l'œuvre ; un doute d'autant plus plausible que le projet n'a pas d'impact sur les indicateurs de santé (poids et taille par âge), résultats intermédiaires par lesquels la vermifugation aurait en toute logique dû transiter pour avoir un effet sur l'absentéisme des enfants. Les épidémiologues critiquent plus généralement la défaillance du protocole de randomisation et de l'analyse subséquente à souscrire aux standards scientifiques codifiés et imposés dans le champ des RCT dans le domaine de la santé publique.

Sans décrire les développements ultérieurs de cette polémique (voir Humphrey, 2015), nous retiendrons de cet exemple, non seulement les prérequis à la généralisation du programme ne sont pas remplis (validité externe), mais aussi que la validité interne de l'étude initiale n'est elle-même pas garantie. Le piquant de cette controverse est que le fer qui est porté contre les *randomistas* provient d'autres randomisateurs (épidémiologistes), reconnus comme les pères fondateurs de la méthode : en quelque sorte l'*evidence based medicine* contre l'*evidence based policy*. Dans ce contexte, il convient de s'interroger sur les facteurs qui ont pu pousser à la généralisation du programme, une extension d'autant plus hasardeuse que les résultats obtenus étaient fragiles. C'est ce à quoi s'attèle la dernière partie du présent article, en élargissant la perspective d'économie politique.

### III. Éléments d'économie politique d'une entreprise scientifique

Comprendre les décalages entre les limites de la méthode et sa très grande légitimité, aussi bien dans le champ académique que politique, suppose de s'interroger sur les rapports de force qui sont en jeu et qui contribuent à forger les préférences collectives en faveur d'une méthode ou d'une autre. L'évaluation d'impact, dont les RCT constituent l'idéal-type, s'est à ce point massifiée qu'il convient aussi de l'appréhender comme une véritable industrie. Comme toute industrie, le marché des évaluations d'impact est la rencontre d'une offre et d'une demande. Cette demande est double : elle provient à la fois de la communauté des donateurs et du monde académique. Quant à l'offre, elle est largement façonnée par un type d'entrepreneurs et d'entreprises scientifiques que nous nous employons à décrire dans cette partie, ainsi que les stratégies qu'ils déploient pour « tenir » le marché.

#### 3.1. Un nouveau scientifique *business model*

Du côté des bailleurs de fonds, la seconde moitié des années 1990 et la décennie 2000 marquent la « fin des idéologies », caractéristique de l'ère des ajustements structurels. La fin de la guerre froide a favorisé une émancipation relative de l'aide publique au développement à l'égard du politique. Pendant cette période, la coopération technique et financière ne constituait souvent qu'un registre supplémentaire des rivalités entre blocs. Cette subordination de la coopération à la *realpolitik* a toutefois été battue en brèche après la chute du Mur de Berlin. Avec la « fin des grands récits », la crise de l'aide, les OMD et le *New Public Management* ont sommé les promoteurs de l'APD d'apporter la preuve de leur utilité (Naudet, 2006).

Le nouveau credo conjugue une focalisation des politiques de développement en faveur de la lutte contre la pauvreté et la mise en avant d'une gestion axée sur les résultats. Ces orientations, formulées dans la *Déclaration de Paris* en 2005, ont été depuis systématiquement réitérées lors de grandes conférences internationales sur l'aide publique au développement à Accra en 2008 puis à Busan en 2011 et enfin à Addis Abeba en 2015. La montée en puissance du paradigme de l'*evidence based policies*, qui consiste à fonder toute décision publique sur des preuves scientifiques, réserve aux savants une légitimité nouvelle dans ces arènes politiques. Les RCT répondent en principe à toutes les conditions requises par ce tournant : empirisme agnostique, simplicité apparente (simple comparaison de moyennes), mobilisation élégante de la théorie mathématique (gage de scientificité), et concentration sur les pauvres (enquêtes auprès des ménages). Leur simplicité les rend aisément compréhensibles par les décideurs. Elles apparaissent donc comme un vecteur privilégié pour éclairer la décision publique. L'évaluation du programme *Progresa* au Mexique a constitué un prototype de cette méthode et un cas d'école de sa performativité, nous l'avons vu plus haut.

Du côté académique, et au premier chef de l'économie, la conjoncture est également favorable à la montée en puissance des RCT : défaite des écoles hétérodoxes centrées sur les structures sociales et les processus de domination, recherche des fondements micro de la macro, primat de la quantification et de l'économie dans le champ des sciences sociales et alignement sur les standards en vigueur au Nord (Milonakis et Fine, 2009). La montée en puissance conjointe de l'économie comportementale et expérimentale, consacrée par l'attribution en 2002 du Nobel d'économie au psychologue Daniel Kahneman et à l'économiste Vernon Smith, spécialistes respectifs de chaque champ, illustre cette évolution de la discipline. Or les RCT se nourrissent et nourrissent en retour de cette montée en puissance, qui est par ailleurs vivement contestée<sup>16</sup>. L'économie expérimentale est une méthode de production de données contrôlées et reproductibles, avec ensuite plusieurs variantes en fonction du degré de contrôle exercé, depuis l'expérimentation en laboratoire jusqu'aux expérimentations naturelles. Les RCT offrent l'opportunité d'expérimentations *sur le terrain*, élargissant ainsi le spectre d'application à divers objets ne se prêtant pas (ou mal) aux expérimentations de laboratoire (List and Metcalfe, 2014). L'économie comportementale, largement fondée sur l'expérimentation, mais pas seulement, se définit par son objet : l'analyse des biais cognitifs, émotionnels et sociaux des comportements individuels. Si elle critique la dimension descriptive de l'hypothèse de rationalité de l'économie néoclassique, elle en conserve la dimension normative, en prescrivant des outils – notamment les *nudges* – supposés corriger les imperfections de comportements. Les RCT mobilisent très largement les préceptes de l'économie comportementale<sup>17</sup>, et c'est d'ailleurs par leur intermédiaire que celle-ci s'est diffusée en économie du développement, jusqu'à y occuper aujourd'hui une place prépondérante (Fine et Santos, 2016).

La Banque mondiale, un acteur majeur doté d'une double légitimité de bailleur financier et académique, a également joué un rôle catalyseur à la fois dans la montée en puissance du

---

<sup>16</sup> Pour un aperçu des controverses sur les expérimentations de terrain, voir Teele (2014). Pour l'économie comportementale, voir Koster et alii (2015).

<sup>17</sup> Les ouvrages de Banerjee et Duflo (2011) et de Karlan et Appel sont à cet égard très illustratifs (2012).

paradigme de l'*evidence based policies* et des RCT. Elle a tout d'abord été le théâtre d'un revirement scientifique, des études classiques en (macro) économie du développement dont le département de la recherche de la Banque était le sérail, vers des nouvelles approches empiriques et tournées vers la microéconomie. Ce retournement a été amorcé en 2003 avec la nomination de François Bourguignon au poste d'économiste en chef. En 2005, celui-ci a contribué à la création d'une unité entièrement consacrée à l'évaluation d'impact (le DIME), financée sur les fonds du département de la recherche. Il a aussi commandité une évaluation des activités de ce dernier. Celle-ci a étrillée les travaux scientifiques menés par la Banque au cours de la décennie précédente, au motif qu'ils étaient avant tout « *utilisés à des fins prosélytes en faveur des politiques de la Banque, sans avoir été soumis à un regard suffisamment impartial et sceptique [et] une carence criante des garde-fous qui commandent de dissocier la recherche et le plaidoyer.* » (Banerjee et alii, 2006, p. 6).

Cette critique fut relayée par un rapport d'un groupe de travail international sur le « fossé de l'évaluation », rassemblant de nombreux chercheurs de renom, dont les principaux promoteurs des RCT (F. Bourguignon, A. Banerjee, E. Duflo, D. Levine, etc.), ainsi que des responsables des principales institutions de développement (CAD, Banque mondiale, Bill et Melinda Gates Foundation, Banques africaine et interaméricaine de développement...). Intitulé « *When will we ever learn* », et publié sous la forme d'un appel-programme par le *Center for Global Development* (Savedoff et alii, 2006), ce rapport a eu un très large écho tant dans la communauté scientifique que parmi les praticiens et les décideurs politiques. Au-delà de son argumentation, ce rapport s'apparente aussi à un plaidoyer *pro domo*, puisqu'il a eu pour effet d'accroître la visibilité et la demande pour les travaux de nombre de ses auteurs, et au premier chef les méthodes expérimentales.

Le Rapport sur le Développement dans le Monde 2015 marque la convergence la plus aboutie des deux mouvements : en faveur des RCT comme méthodologie en général, d'une part ; en faveur de l'économie comportementale comme angle d'approche disciplinaire, de l'autre. Il entend reformuler les politiques de développement sur la base d'une « meilleure compréhension du comportement humain » et le recours aux nudges (définis dans le Rapport comme « *a policy that achieves behavior change without actually changing the set of choices* » ; Banque mondiale, 2015, p. 36) pour corriger les imperfections de comportement, est emblématique de l'enchevêtrement entre RCT, économie comportementale et expérimentale, mentionné plus haut (Fine et Santos, 2016).

Mais l'élan en faveur des RCT a surtout été porté par l'émergence d'une nouvelle génération de chercheurs. Ils sont jeunes, issus du sérail des meilleures universités (pour la plupart américaines)<sup>18</sup>. Ils ont su trouver la formule du carré magique en combinant excellence académique (légitimité scientifique), effort de séduction en direction du public (visibilité médiatique et légitimité citoyenne) et des bailleurs de fonds (demande solvable), investissement massif dans la formation (offre qualifiée), et modèle d'entreprise performant (rentabilité financière) ; toutes ces qualités se renforçant mutuellement. En multipliant des cours dans les cursus universitaires, mais également en proposant des sessions courtes de formation destinées à

---

<sup>18</sup> A. Jatteau (2016) montre que les chercheurs de J-PAL sont à la fois plus souvent issus d'établissement d'élites et occupent en moyenne des postes plus prestigieux que leurs homologues (détenteurs de chaires, Ivy League Plus, NBER, BREAD, CEPR, etc.).

un large public, en assurant des enseignements classiques (en présentiel) mais également sous des formes nouvelles (MOOC), les randomisateurs se donnent les moyens d'attirer des ressources jeunes, motivées et hautement qualifiées<sup>19</sup>. En s'engageant dans une intense activité de communication et de plaidoyer, à l'aide de toute une série de supports de presse ou para-académiques (*policy briefcases*, blogs, forums de vulgarisation, *happenings*, etc.), ils montrent l'image avenante de chercheurs ayant accepté de sortir de leur tour d'ivoire. En affichant une posture modeste au plus proche du terrain<sup>20</sup>, ils incarnent l'engagement, l'empathie et le désintéressement.

L'étude de la trajectoire et des élites et de leurs réseaux s'avère une perspective puissante pour comprendre l'émergence, le déclin ou la diffusion transnationale de paradigmes scientifiques ou politiques (Dezalay et Garth, 2002). Sans prétendre mener cette entreprise de manière exhaustive à propos des RCT, qui constituerait un programme de recherche à soi seul, nous nous contenterons d'en tracer les linéaments. L'analyse menée par A. Jatteau (2016, pp. 305-334) sur les réseaux des *randomistas* et leur connexité, à partir du corpus de chercheurs impliqués dans la conduite de RCT et des auteurs de publications académiques associées, met en évidence l'existence d'un réseau d'interconnaissance à la fois dense et hiérarchique, avec l'émergence de deux leaders (des « nœuds » en théorie des réseaux) : Dean Karlan et Esther Duflo, cette dernière étant la figure emblématique du mouvement. Jeune chercheuse franco-américaine, E. Duflo cumule les distinctions académiques, dont la célèbre médaille Bates, qui récompense le "meilleur économiste" de moins de 40 ans et qu'elle a reçue en 2010. Elle a à son actif un nombre impressionnant de publications dans les revues d'économie les plus prestigieuses. Mais elle vulgarise également ses travaux sous la forme d'ouvrages accessibles au grand public et succès de librairie (voir par exemple Banerjee et Duflo, 2011 ; en français : Duflo, 2010). Depuis 2008, elle figure sur la liste des 100 premiers intellectuels mondiaux du magazine américain *Foreign Policy*. En 2011, c'est le *Time* qui la compte parmi les 100 personnes les plus influentes au monde. Fin 2012, elle a été nommée conseillère du Président Obama sur les questions de « développement global ». En France, elle a été la première titulaire de la toute nouvelle chaire « Savoir contre pauvreté » du Collège de France, financée par l'AFD (Agence Française de Développement). En 2015, elle obtient le *Prix Princesse des Asturies* des Sciences sociales, une prestigieuse distinction espagnole. Son nom est régulièrement cité comme candidate potentielle à un prochain prix Nobel d'économie.

Ces jeunes chercheurs de la mouvance RCT se distinguent également dans le mode de gestion de leur activité. En montant des ONG ou des bureaux d'étude spécialisés, ils créent les structures idoines pour recevoir des fonds de toutes origines : publique bien sûr, mais également de fondations, d'entreprises, de mécènes, etc., hors des circuits classiques de financement de la recherche publique. Sur ce plan, ils sont en parfaite adéquation avec les nouvelles sources de

---

<sup>19</sup> En témoigne le succès des formations en ligne spécifiquement dédiées à ce type de méthode, dispensées sur la plateforme eDX du MIT ; voir par exemple le MOOC intitulé [The Challenges of Global Poverty](#) et dispensé par Banerjee et Duflo ou encore celui intitulé [Evaluating Social Programs](#), par deux de leurs collègues de J-PAL, que plus de 1 000 participants ont suivi en 2013.

<sup>20</sup> On a vu plus haut qu'étant donné le nombre de RCT entreprises simultanément, leur connaissance du terrain posait question.

financement de l'aide que constituent les fondations privées et les institutions philanthropiques, qui se montrent particulièrement enclines à leur confier des études. En parvenant à créer leurs propres guichets de financement, principalement multilatéraux (l'initiative de la Banque mondiale pour l'évaluation d'impact du développement, l'initiative internationale pour l'évaluation d'impact, la Banque africaine de développement, le Fonds Stratégique pour l'Evaluation d'Impact (SIEF)), mais aussi bilatéraux (la coopération espagnole et britannique) ainsi qu'en provenance de grandes fondations (Rockefeller, Citi, Gates, MacArthur, Hewlett), les randomisateurs ont créé un oligopole sur le marché florissant de l'expérience aléatoire, même si la concurrence est aujourd'hui plus vive compte tenu de l'adoption des méthodes RCT par un nombre grandissant d'équipes de recherche. La nébuleuse organisée autour de J-PAL constitue le modèle le plus emblématique et le plus abouti de ce nouveau *scientific business model*. Le laboratoire J-PAL est rattaché au département d'économie du *Massachusetts Institute of Technology*. Cet ancrage institutionnel, au sein d'une des plus prestigieuses universités américaines, ainsi que la notoriété de ses dirigeants, sert à la fois de caution académique et de catalyseur.

Au côté de J-PAL, Innovations for Poverty Action (IPA) joue un rôle névralgique. Cette organisation à but non lucratif a réussi à générer plus de 250 millions de dollars de revenus depuis 2003, date de sa création, des montants annuels qui ne cessent de croître (Jatteau, 2016, p. 265). Outre sa fonction de communication et de plaidoyer en faveur des RCT<sup>21</sup>, elle est chargée d'étendre et de répliquer les expériences aléatoires une fois testées par J-PAL. C'est donc l'articulation des deux institutions qui doit permettre de mener à bien le « projet généralisateur » de réplification décrit dans la deuxième partie. Annie Duflo, la sœur d'Esther Duflo, en est le directeur exécutif. Dean Karlan (mentionné plus haut pour ses 100 RCT), professeur à Yale et ex-doctorant des deux initiateurs de J-PAL, est lui le fondateur et membre du bureau des directeurs. Comme Abijit Banerjee est également le conjoint d'Esther Duflo, J-PAL/IPA est non seulement une entreprise globale mais également une affaire de famille. Plus largement, les frontières entre les deux institutions sont poreuses, et de nombreux membres et associés y exercent des responsabilités croisées.

L'industrie des RCT est une entreprise rentable sous tous rapports. Elle est académiquement profitable, et il y a tout à gagner à s'inscrire dans cette mouvance (ou tout à perdre à ne pas en être). Il est aujourd'hui très difficile de publier dans des revues d'économie des articles basés sur d'autres approches. Cet effet d'éviction est aussi lié au fait que les promoteurs les plus influents des RCT sont souvent membres des comités de rédaction des plus grandes revues d'économie et d'économie du développement<sup>22</sup>. Le numéro spécial de l'*American Economic Journal: Applied*

---

<sup>21</sup> A la différence de J-PAL, contraint par les réserves académiques de rigueur, IPA peut afficher ses objectifs en usant des recettes directement issues du marketing. Par exemple, "IPA uses randomized evaluations because they provide the highest quality and most reliable answers to what works and what does not" (voir le site d'IPA : <http://www.poverty-action.org/>).

<sup>22</sup> Comme par exemple et parmi les plus fameuses : *Annual Review of Economics*, *Journal of Economic Literature*, *The American Economic Journal: Applied Economics*, *Review of Economics and Statistics*, *Journal of Development Economics*, *Review of Development Studies*, *Journal of Quantitative Economics*, *Journal of Economic Perspectives*, etc. Encore une fois D. Karlan apparaît comme le plus boulimique : il est membre dans au moins huit comités

*Economics* consacré aux RCT dans le domaine du microcrédit est illustratif à cet égard. Les trois éditeurs scientifiques du dossier sont membres de J-PAL. En dehors de l'introduction générale, chacun co-signe un article, et deux d'entre eux sont membres du comité éditorial (Banerjee et Karlan). Ester Duflo est à la fois Editeure (fondatrice) de la revue et co-auteure de deux des six articles. Si on ajoute que près de la moitié des auteurs/articles (soit 11 sur 25) sont également membres de J-PAL, et quatre autres sont chercheurs associés ou doctorants, on s'éloigne des principes de contrôle par les pairs censés régir la publication scientifique. On retrouve sur cet exemple ponctuel, l'image réduite de l'exceptionnelle densité des liens entre *randomistas* mise en lumière par Jatteau (2016).

Mais les profits ne sont pas uniquement symboliques. Se spécialiser dans les RCT est aussi une excellente façon d'obtenir un poste de chercheur ou d'enseignant, comme en témoignent les modes de recrutement actuels dans la discipline économique. C'est aussi une garantie d'obtention de fonds conséquents pour mener ses propres recherches (dans un contexte de pénurie généralisée) et de générer des rémunérations additionnelles très substantielles, sous forme de consultations ou de participations à des instances de direction (Jatteau, 2016).

### 3.2. Les stratégies de contrôle du marché et de capture des rentes

Dans ce contexte, on comprend mieux pourquoi les critiques adressées aux méthodes expérimentales sont particulièrement mal accueillies et rencontrent une très forte résistance de la part de leurs promoteurs. Plusieurs stratégies sont mises en œuvre pour asseoir le monopole et éviter un rééquilibrage au tout-RCT. Comme ce fut le cas en médecine, les méthodes alternatives sont disqualifiées, les RCT s'arrogeant le monopole de la scientificité (Harrison, 2011). La référence à l'*evidence-based medicine* est mobilisée comme gage de scientificité, mais en occultant les multiples controverses dont elle a fait l'objet, et qui se poursuivent aujourd'hui (Jacobson *et alii*, 1997 ; Schulz *et alii*, 1995 ; Labrousse, 2010). Le débat en face à face est souvent esquivé ou refusé<sup>23</sup>. Les réflexions critiques sont longtemps restées inaudibles, car cantonnées sur des supports de publication marginalisés. Dans de nombreux cas, les résultats des expériences présentés comme des « découvertes » inédites, ne sont en fait que la reprise de conclusions obtenues par des études antérieures. Le subterfuge procède en deux temps. D'abord, l'essentiel de la littérature existante est discréditée, puisqu'elle est considérée a priori comme non rigoureuse, les RCT étant revendiquées supérieures et appréhendées comme le seul mode définitif d'administration de la preuve. Ce dénigrement presque rituel conduit à faire des connaissances accumulées dans le passé un champ vierge (*reset*), qui permet, dans un second temps, de faire passer tout résultat issu de RCT comme une « découverte » majeure, malgré son caractère

---

(board), soit cinq des revues citées plus haut, auxquelles il faut ajouter *Behavioral Science & Policy*, *Stanford Social Innovation Review* et *Journal of Globalization and Development*.

<sup>23</sup> Voir à ce sujet l'entretien d'A. Deaton avec A. Jatteau (2016, p. 60). Le parcours scientifique d'A. Jatteau, décrite dans sa thèse, mis à l'écart du réseau J-PAL après en avoir fait partie pour visée critique est lui-même riche d'enseignements. Plus modestement et à notre échelle, toutes les tentatives d'engager un débat contradictoire avec les promoteurs du mouvement (par voie de publications ou lors de conférences) ont été éludées.



éventuellement (et souvent) redondant<sup>24</sup>. La manipulation est d'autant plus facile à faire admettre que les publications tirées de méthodes non expérimentales ne sont presque jamais citées (Labrousse, 2010 ; Nubukpo, 2012).

La prise de pouvoir par les randomisateurs n'est donc pas seulement une question de méthode. Elle s'étend au contenu et aux thèses qui ont cours sur les grandes questions scientifiques et de politique dans le champ du développement. Outre le cas emblématique des vermifuges, mentionné plus haut, quelques exemples pris dans le cadre de la microfinance sont de ce point de vue illustratifs de cette propension à faire passer pour originales, des thèses déjà largement explorées dans la littérature (Bédécarrats *et alii*, 2015)<sup>25</sup>.

Le décalage entre la littérature et la pratique reflète bien la relative autonomie de la recherche *mainstream* par rapport à la pratique et son étanchéité par rapport à des études pourtant robustes menées par ailleurs. La littérature académique publiée dans de prestigieuses revues internationales s'est éprise de la soi-disant révolution financière du prêt de groupe, alors qu'elle était pratiquée par nombre d'institutions de microfinance depuis les années 1960 (Gentil, 1996 ; Dichter et Harper, 2007). La question de l'efficacité du prêt collectif a été tranchée de manière très pragmatique par les acteurs de terrains qui ont dès le milieu des années 1990 développé le crédit individuel (à commencer par la Grameen Bank). Un pan de la recherche plus opérationnelle a montré dès le début des années 2000 que le prêt de groupe n'est justifié que dans certains contextes qui s'y prêtent socialement et économiquement (réseaux sociaux denses mais pas trop hiérarchiques, faible spécialisation des emprunteurs dans le même secteur, etc.). D'autres travaux, de nature académique, et sur la base d'études quantitatives robustes parviennent à des conclusions similaires (Godquin, 2004 ; Sharma et Zeller, 1997 ; Gonzalez-Vega *et alii*, 1996). Or les travaux expérimentaux ignorent complètement ces différents pans de littérature et n'apportent rien de nouveau, hormis le mode d'administration de la preuve. Par exemple les études de Giné et Karlan (2014) sur les effets de la garantie solidaire (versus individuelle) sur les impayés sont redondantes par rapport aux dizaines d'études déjà consacrées au sujet et notamment aux trois citées plus haut. Plus largement, la conclusion de randomisateurs selon laquelle « *le microcrédit n'est peut-être pas un "miracle", comme on l'a parfois prétendu, mais il permet effectivement aux ménages d'emprunter, d'investir, et de créer ou d'étendre leurs activités* » (Banerjee *et alii*, 2015a), n'est pas très novatrice. On la retrouve dans de nombreuses études d'impact parmi les 170 publiées sur la microfinance (dont une dizaine de RCT) entre 1980 et 2010 (Bédécarrats, 2012).

Les exemples pourraient être multipliés à l'envi. Qu'il s'agisse de la caution solidaire (versus caution individuelle), ou de l'épargne, la littérature préexistante aux RCT interroge le caractère novateur des enseignements de ces dernières. Elles arrivent plus d'une quinzaine d'années après

---

<sup>24</sup> Ce mode opératoire bien connu a déjà été utilisé en économie du développement à propos des institutions. Après avoir exclu du champ de la discipline légitime les courants hétérodoxes (écoles de la régulation et des conventions, néo-institutionnalisme, etc.) dont c'était une des originalités, l'économie *mainstream* s'est réappropriée ce sujet, en faisant passer ses résultats comme des nouveautés. Le même phénomène est actuellement à l'œuvre avec l'économie politique.

<sup>25</sup> Les thèses développées par Banerjee et Duflo (2011) dans leur ouvrage *Poor Economics* méritaient à elles seules d'être passées à ce crible.

que les innovations financières aient été expérimentées, documentées et largement diffusées. Ce constat contredit le message véhiculé par J-PAL et IPA, qui se présentent avant tout comme des dispositifs de tests stimulant l'innovation en matière de politiques sociales<sup>26</sup>.

Alors que tout concourrait à ce que s'engage une véritable controverse scientifique (au sens de Callon *et alii*, 2001 ; Knorr-Cetina, 1982) sur la question des expériences aléatoires, celle-ci n'a toujours pas vu le jour du fait de l'asymétrie de pouvoir des acteurs en présence. Néanmoins, il convient de noter que la multiplication des critiques conduit lentement à faire bouger les lignes, comme le montrent l'évolution de la dynamique du champ scientifique retracée par nos travaux successifs sur la question (Bédécarrats *et alii*, 2013, 2015 et le présent article). Toutes les critiques, qui en tout état de cause restent minoritaires, n'ont pas la même portée ni le même écho. De ce point de vue, la persistance d'Angus Deaton à s'en prendre au projet hégémonique des *randomistas*, qui après son article de 2010, s'associe à Nancy Cartwright, philosophe des sciences de renommée internationale et spécialiste des RCT dans le champ de la médecine, pour publier un nouvel article sévèrement critique, est plus difficile à esquiver, d'autant qu'il a été nobélisé entre temps<sup>27</sup>.

Encore une fois le numéro spécial consacré au microcrédit peut être cité en exemple. Ainsi, le numéro est fourni avec les bases de données originales, pour répondre au grief d'opacité et faciliter les méta-analyses. L'introduction générale synthétise les réponses apportées (Banerjee *et alii*, 2015a). Un modèle théorique est développé, pour répondre aux critiques d'empirisme agnostique. Les questions du taux de *take up*, de précision des estimateurs ou de l'hétérogénéité du traitement sont reconnues (validité interne). La diversité contextuelle est prise en compte à travers la diversité des terrains, des produits et des institutions abordées dans les six articles (validité externe). L'analyse coûts-bénéfices proposée par Banerjee *et alii* (2015a) est censée répondre à la question de l'efficacité. Mais le changement de posture reste marginal. En particulier, les deux principes fondateurs demeurent : les RCT constituent la méthode la plus rigoureuse pour mesurer l'impact causal ; le projet généralisateur visant à établir ce qui marche et ce qui ne marche pas est maintenu inchangé. Ainsi, l'argument sur la prise en compte de la diversité des contextes est retourné. La similitude des résultats obtenus dans les six pays, considérés comme un « *échantillon raisonnablement représentatif de l'industrie/mouvement du microcrédit à l'échelle mondiale* » (Banerjee *et alii*, 2015a, p. 2) permettrait de lever la critique portée sur la validité externe de la méthode.

Cet évitement de la controverse est d'autant plus problématique que l'objectif ultime affiché des RCT – peser sur les politiques publiques – est largement sujet à caution. Sur les 543 RCT menées par J-Pal et terminées (parmi les 811 qui ont été entreprises), 11 ont fait l'objet d'un *scaling up*, soit 2%. Encore convient-il de noter que 5 d'entre-elles concernent le « *deworming* », dont nous avons montré les failles dans la partie II ; et 3 autres la distribution de chlore, dont l'apport a été

---

<sup>26</sup> Cf. la description qu'IPA donne de sa stratégie, à la page : <http://www.poverty-action.org/about>, consultée le 22 janvier 2015.

<sup>27</sup> Sans se livrer à l'exégèse de l'article de 2016, on notera qu'il développe toute une batterie de nouvelles critiques, au point que le papier de 2010 n'est même pas cité en référence du second.

sérieusement nuancé par le LSHTM (Wolf *et alii*, 2014). Ce point mériterait une analyse à part entière, mais trois arguments peuvent néanmoins être avancés. Le premier est la complexité des prises de décision en matière de politiques publiques, qui explique d'ailleurs que les méthodes expérimentales, quel que soit leur champ d'application, ont en fait un effet *direct* limité<sup>28</sup>. Le deuxième argument, déjà évoqué, concerne l'étroitesse des questions susceptibles de se prêter aux canons de la randomisation, ce qui laisse de côté une très large partie des questions de développement (cf. *supra*, 3.1.). Lorsque la Banque mondiale (2015) entend redéfinir les politiques de développement, les propositions sont par exemple des tirelires en métal pour augmenter l'épargne de ménages pauvres, des spots télévisés pour réduire la consommation d'eau, des autocollants dans des bus pour limiter les accidents de la route, la livraison de fertilisants à une période appropriée pour améliorer les conditions de vie des petits agriculteurs. Accepter l'idée que les RCT pourraient avoir un impact sur les politiques de développement suppose d'admettre que celui-ci se résume désormais à accompagner les populations à réaliser de meilleurs choix, mais dans un environnement qui demeure inchangé.

Enfin, le troisième argument est lié aux objectifs et aux contraintes de publication académique, dictés par la pression considérable exercée sur la profession en matière de publication. Il pèse fortement sur la mise en œuvre des RCT. C'est ce que met en évidence Arthur Jatteau (2016) au travers d'une analyse exhaustive de la production de J-PAL et de multiples entretiens avec des *randomistas* et de leurs assistants de recherche. Qu'il s'agisse du choix du sujet, de l'échantillonnage, des modules de questionnaire ou encore du type de résultats mis en avant, les choix opérés visent bien souvent à maximiser le rapport entre moyens humains et financiers investis et le nombre de *publishable units* qu'on peut en tirer. Les revues académiques sont avant tout intéressées par des projets novateurs (quels qu'ils soient), et non par la consolidation de connaissances déjà publiées dans de nouveaux contextes. Ces arbitrages s'effectuent souvent au détriment des attentes des partenaires opérationnels de l'expérimentation, de la pertinence pour les populations ou de l'utilité pour la décision politique. Nombre de RCT sont déconnectées des réalités locales (Barrett et Carter, 2010), mais vendeuses académiquement parlant (Jatteau, 2016, p. 423-432), avec deux biais particulièrement fréquents : une focalisation excessive sur des projets inédits et supposés novateurs, et sur des résultats positifs (alors que les effets négatifs où nuls seraient tout autant instructifs, mais souvent passés sous silence). Le hiatus entre logique académique portée par les *randomistas* les plus en vue et la logique économique et de politique publique, que tente de dénouer l'association entre JPAL et IPA, risque à terme de déboucher sur une impasse, à mesure que les attentes des financeurs de RCT en termes de passage à l'échelle, attentes elles-mêmes entretenues par les *randomistas*, seront déçues.

## Conclusion

Cet article s'est employé à décrire la montée en puissance des méthodes expérimentales dans le champ du développement, au point de devenir l'étalon-or de l'évaluation d'impact. Il a ensuite montré les limites méthodologiques des RCT et l'inanité du projet hégémonique caressé par leurs

---

<sup>28</sup> Voir par exemple Moffitt (2014) à propos des *Social Welfare Programs*.

promoteurs. Enfin, il a proposé quelques éléments d'interprétation pour comprendre les facteurs qui ont présidé à l'instauration de cette nouvelle norme internationale, dans une perspective d'économie politique. De notre point de vue, les expériences aléatoires appliquées au développement représentent un progrès méthodologique. Néanmoins, cette avancée circonscrite s'accompagne d'une double régression : épistémologique d'abord, les promoteurs des RCT partageant une conception positiviste de la science, aujourd'hui surannée ; politiquement ensuite, par le caractère impérialiste d'une démarche prétendant comprendre tous les mécanismes de développement par cet instrument. De plus et ça n'est pas le moindre des paradoxes du succès des randomistas que d'avoir réussi à labéliser la méthode comme la seule à même d'identifier rigoureusement des impacts causaux, alors que les RCT ne donnent en fait que des preuves d'efficacité sans rien dire des mécanismes causaux à l'œuvre.

Parmi les extensions possibles de cette recherche, deux pistes paraissent prometteuses : l'une analytique et l'autre méthodologique. Sur le premier front, notre approche d'économie politique méritait d'être complétée par des travaux s'inscrivant dans le registre de l'histoire ou de la sociologie des sciences. Dans une perspective « latourienne », des recherches portant sur les interactions entre la production scientifique, et les conditions sociales, la personnalité des acteurs ou encore l'architecture institutionnelle pourraient s'appliquer fructueusement à l'industrie des RCT, à ses figures tutélaires et à ses centres de recherche les plus en vue : l'intérêt pour la vie de laboratoire ne devrait pas être seulement réservé aux sciences dites « dures » (Latour, 1999). On peut aussi penser à des approches historiques de trajectoire de randomisateurs, à l'instar de ce qui se fait pour des capitaines d'industrie, des hommes politiques ou encore des scientifiques prestigieux. Les analyses du laboratoire J-PAL par Jatteau (2016), si elles sont loin d'épuiser le sujet, en montrent la fécondité.

Sur le second front, notre propos n'est pas de rejeter les RCT, qui constituent une méthode prometteuse... parmi d'autres ; encore convient-il de les mener dans les règles de l'art, et un alignement sur les bonnes pratiques instaurées dans le domaine médical est une nécessité. Si les RCT restent probablement adaptées et légitimes pour certaines politiques circonscrites de manière précise, il est à la fois nécessaire et possible d'employer d'autres méthodes. Celles-ci adoptent une approche pragmatique, définissant les questions de recherche et les outils méthodologiques nécessaires au cas par cas, en concertation avec les partenaires impliqués (opérateurs de terrain, bailleurs, etc.). Elles s'appuient également sur un pluralisme méthodologique, fondé sur l'interdisciplinarité et reconnaissent la diversité des modes d'administration de la preuve (inférence statistique/analyse compréhensive). Il ne s'agit pas de récuser le formalisme ou la modélisation, mais d'en faire un usage contrôlé. En outre, ces approches ne visent pas l'énoncé de lois universelles, mais cherchent à expliciter des liens de causalité propres à une période et à un contexte précis. Les méthodes qualitatives sont mobilisées pour contextualiser les politiques de développement, élaborer des hypothèses originales, identifier des phénomènes nouveaux ou imprévus, et enfin les analyser dans leur globalité en étudiant la complexité des liens de causalité, les interactions multiples, dynamiques et contradictoires entre différentes entités, de manière située. L'idée d'*iterative learning*, suggérée par Pritchett *et alii* (2013),

et suggérant un design d'étude dynamique, issu d'interactions permanentes entre les résultats obtenus et les modalités du projet, est particulièrement stimulante.

Les degrés d'articulation et d'intégration entre méthodes (quantitatives, qualitatives et participatives) peuvent être très variés. Plutôt que de miser systématiquement sur la création de données nouvelles, ces méthodes alternatives privilégient, lorsque cela se justifie, l'usage de données existantes, qu'il s'agisse de statistiques officielles ou de données produites par les partenaires locaux en charge de la mise en œuvre des projets/politiques de développement. Cela permet notamment de réduire les coûts et la lourdeur des protocoles d'enquêtes, dont on a vu plus haut les incidences potentiellement néfastes sur la qualité des recherches menées. Cela permet aussi de renforcer les systèmes d'information des programmes évalués ou les appareils statistiques des pays dans lesquelles ils se situent. Après avoir été emportés par l'effet de mode, certains bailleurs de fonds comme l'AFD (2013) et plus tièdement le DFID (2012) se montrent aujourd'hui plus circonspects quant au champ réel d'application des RCT et à leur capacité à répondre aux questions qu'ils se posent. Il reste à espérer que ce retour à des positions plus nuancées se traduise par des mesures concrètes de soutien à des méthodes d'évaluation alternatives et / ou complémentaires.

Déjà en 2012, R. Picciotto (2012) se demandait quand la bulle (des RCT) allait éclater. Les attaques sont grandissantes, et la nobélisation d'A. Deaton, fervent critique des RCT, va probablement accélérer le processus. Ce qui apparaît clairement, c'est que le mouvement des RCT fera tout pour éviter le déclassement afin de continuer à bénéficier des retombées de la position dominante qu'il occupe dans le champ de l'évaluation d'impact. Il est possible que les *randomistas* finissent par reconnaître que leurs prétentions à vouloir trancher ce qui marche et ce qui ne marche pas dans tous les domaines étaient un leurre, mais ce retournement n'interviendra que sous la pression et une fois le maximum de bénéfices engrangés<sup>29</sup>. Enfin, pour que la bulle éclate, il faudrait aussi des alternatives crédibles, ce qui implique de poursuivre activement la réflexion et l'action en matière d'innovation méthodologique.

---

<sup>29</sup> On peut d'ailleurs s'interroger sur la sincérité des membres du mouvement à défendre leur projet face aux critiques. Il est probable que certains continuent d'y croire, tandis que d'autres sont conscients de ses limites, mais, plus cyniques, souhaitent profiter des privilèges et avantages qu'il procure encore. Les entretiens menés par Jatteau (2016) offre une palette de ces motivations.

## References

Acemoglu D. (2010), “Theory, general equilibrium, and political economy in development economics” *Journal of Economic Perspectives* 24 (3): 17-32

AFD (2013), *Stratégie 2013-2016 en matière d'évaluations d'impact*. Paris: AFD.

Aiken A. M., C. Davey, J.R. Hargreaves and R.J. Hayes (2015), “Re-analysis of health and educational impacts of a school-based deworming programme in western Kenya: a pure replication” *International Journal of Epidemiology* 44 (5): 1572-1580.

Angrist J.D. and J.S. Pischke (2009), *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Banerjee A., A. Deaton, N. Lustig and K. Rogoff (2006), *An Evaluation of World Bank Research, 1998-2005*. Washington D.C.: World Bank.

Banerjee A. and E. Duflo (2011), *Poor Economics: a Radical Rethinking of the Way to Fight Global Poverty*. New York: Public Affairs.

Banerjee A., E. Duflo, R. Glennerster and C. Kinnan (2015a), “The Miracle of Microfinance? Evidence from a Randomized Evaluation” *American Economic Journal: Applied Economics* 7 (1): 22-53.

Banerjee A., E. Duflo, N. Goldberg, D. Karlan, R. Osei, W. Parienté, J. Shapiro, B. Thuysbaert and C. Udry (2015b), “A multifaceted program causes lasting progress for the very poor: Evidence from six countries” *Science* 348 (6236): 1-16. [DOI: 10.1126/science.1260799]

Banerjee A. and R. He (2008), “Making Aid Work” in: W. Easterly, *Reinventing Foreign Aid*. Boston: MIT Press.

Banerjee A., D. Karlan and J. Zinman (2015a), “Six Randomized Evaluations of Microcredit: Introduction and Further Steps” *American Economic Journal: Applied Economics* 7 (1): 1-21.

World Bank (2015), *World Development Report 2015: Mind, Society, and Behavior*. Washington D.C.: World Bank.

Barrett C.B. and M.R. Carter (2010), “The Power and Pitfalls of Experiments in Development Economics: Some Non-Random Reflections” *Applied Economic Perspectives and Policy* 32 (4): 515-548.

Bédécarrats F., I. Guérin and F. Roubaud (2015), “The gold standard for randomized evaluations: from discussion of method to political economy”, *DLAL*, Working Paper 2015-01.

Bédécarrats F., I. Guérin and F. Roubaud (2013), « L'étalon-or des évaluations randomisées : du discours de la méthode à l'économie politique » *Sociologie pratique* 2013/2 (27): 107-122.

Bédécarrats F. (2012), « L'impact de la microfinance : un enjeu politique au prisme de ses controverses scientifiques » *Mondes en développement* 158: 127-142.

Bernard T., J. Delarue and J.-D. Naudet (2012), “Impact evaluations: a tool for accountability? Lessons from experience at Agence Française de Développement” *Journal of Development Effectiveness* 4 (2): 314-327.

- Bold T., M. Kimenyi, G. Mwabu, A. Nganga, J. Sandefur, R.J. DiClemente, A.L. Swartzendruber, J.L. Brown, M. Medeiros and D. Diniz (2013), “Scaling up what works: Experimental evidence on external validity in Kenyan education”, *Center for Global Development*, Working Paper 321.
- Boseley S. (2015), “New research debunks merits of global deworming programmes”, *The Guardian*, 23 July.
- Cahuc P. and T. Le Barbanchon (2009), “Labor Market Policy Evaluation in Equilibrium: Some Lessons of the Job Search and Matching Model”, *CNRS/Ecole Polytechnique*, Working Paper 2008-40.
- Cahuc P. and A. Zylberberg (2016), *Le négationnisme économique et comment s’en débarrasser*, Paris: Flammarion.
- Callon M., P. Lascoumes and Y. Barthe (2001), *Agir dans un monde incertain. Essai sur la démocratie technique*, Paris: Le Seuil, collection La couleur des idées.
- Cartwright N. (2010), “What are randomised controlled trials good for? ” *Philosophical studies* 147 (1): 59-70.
- Cling J.-P., M. Razafindrakoto and F. Roubaud (Ed.) (2003), *Les nouvelles stratégies internationales de lutte contre la pauvreté*, Paris: Economica/IRD.
- Davey C. (2015), “Re-analysis of health and educational impacts of a school-based deworming programme in western Kenya: a statistical replication of a cluster quasi-randomized stepped-wedge trial” *International Journal of Epidemiology* 44 (5): 1581-1592.
- Deaton A. (2010), “Instruments, Randomization and Learning about Development” *Journal of Economic Literature*, 48 (2): 424-455.
- Deaton A. and N. Cartwright (2016), “Understanding and Misunderstanding Randomized Controlled Trials”, *NBER*, Working Paper 22595.
- Devaradjan S. (2013), “Africa's statistical tragedy” *The Review of Income and Wealth* 59: 1-7.
- Dezalay Y. and B.G. Garth (2002), *The Internationalization of Palace Wars. Lawyers, Economists, and the Contest to Transform Latin American States*, Chicago: Series in Law and Society.
- DFID (2012), “Broadening the Range of Designs and Methods for Impact Evaluations. Report of a Study commissioned by the Department for International Development”, *DFID*, Working Paper 38.
- Dichter T. and M. Harper (Ed.) (2007), *What's Wrong with Microfinance?*, Rugby: Practical Action Publishing.
- Duflo E. (2016), “Randomized Controlled Trials, Development Economics and Policy Making in Developing Countries”, World Bank Conference: The State of Economics, *The State of the World*, Washington D.C., online at: <https://olc.worldbank.org/content/state-economics-influence-randomized-controlled-trials-development-economics-research-and>.
- Duflo E. (2010), *Lutter contre la pauvreté*, Paris: Seuil, collection La République des idées.

- Duflo E., P. Dupas and M. Kremer. (2015), “School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools”, *Journal of Public Economic* 123: 92-110.
- Duflo E., P. Dupas and M. Kremer (2012), “School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools”, *NBER*, Working Paper 17939.
- Duflo E. and M. Kremer (2005), “Use of randomization in the evaluation of development effectiveness”, in: G.K. Pitma, O.N. Feinstein and G.K. Ingram (Ed.), *Evaluating Development Effectiveness*. World Bank Series on Evaluation and Development (7), New Brunswick: Transaction Publishers: 205-231.
- Duflo E., R. Glennerster and M. Kremer (2007), “Using Randomization in Development Economics Research: A Toolkit”, in: T.P. Schultz and J. A. Strauss (Ed.), *Handbook of Development Economics*, Vol. 4, London: Elsevier: 3895–3962.
- Easterly W. (2007), *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good*, Oxford: Oxford University Press.
- Eble A., P. Boone and D. Elbourne (2014), “Risk and evidence of bias in randomized controlled trials in economics”, Mimeo, Brown University.
- Faulkner W. N. (2014), “A critical analysis of a randomized controlled trial evaluation in Mexico: Norm, mistake or exemplar?” *Evaluation* 20 (2): 230-243.
- Fine B. and A. Santos (2016), “Nudging or Fudging: The World Development Report 2015” *Development and Change* 47 (4): 640-663.
- Gentil D. (1996), “Les Avatars du modèle Grameen Bank” *Revue Tiers-Monde* 37 (145): 115-133.
- Giné X., D.S. Karlan (2014), “Group versus individual liability: Short and long term evidence from Philippine microcredit lending groups” *Journal of Development Economics* 107: 65-83.
- Glennerster R. (2012), “The Power of Evidence: Improving the Effectiveness of Government by Investing in More Rigorous Evaluation” *National Institute Economic Review* 219 (1): R4-R14.
- Godquin M. (2004), “Microfinance Repayment Performance in Bangladesh: How to Improve the Allocation of Loans by MFIs” *World Development* 32 (11):1909-1926.
- Gonzalez-Vega C., R.L. Meyer, M. Schreiner, J. Rodriguez-Meza, G.F. Monje (1996), *Microfinance Market Niches and Client Profiles in Bolivia*, Columbus: Ohio State University Press.
- Harrison G.W. (2011), “Randomisation and Its Discontents” *Journal of African Economics* 20 (4): 626-652.
- Heckman J.J. (1991), “Randomization and Social Policy Evaluation”, NBER Technical Working Paper 107.
- Heckman J.J., J. Smith and C. Taber (1998), “Accounting for dropouts in evaluations of social programs” *Review of Economics and Statistics* 80: 1-14.



- Heckman J. J., E. Vytlacil (2005), “Structural equations, treatment effects, and econometric policy evaluation” *Econometrica* 73 (3): 669-738.
- Humphreys M. (2015), “What has been learned from the deworming replications: a nonpartisan view”, Columbia: Columbia University, online at: [www.columbia.edu/~mh2245/w/worms.html](http://www.columbia.edu/~mh2245/w/worms.html).
- IEG (2012), *World Bank Group Impact Evaluation. Relevance and Effectiveness*, Washington D.C.: World Bank.
- Knorr-Cetina K. D. (1982), “Scientific communities or Transepistemic Arenas of Research? A Critique of Quasi-Economic Models of Science” *Social Studies of Science* 12: 101-130.
- Jacobson L., A. Edwards, S. Granier, C. Butler (1997), “Evidence-based Medicine and General Practice” *British Journal of General Practice* 47: 449-452.
- Jatteau A. (2016), *Faire preuve par le chiffre ? Le cas des expérimentations aléatoires en économie*, PhD Dissertation, Université Paris-Saclay, Ecole Normale Supérieure Paris-Saclay.
- Jerven M. (2015), *Africa: Why Economists Get It Wrong, African Arguments*, Zed Books, London.
- Jerven, M. (2013), *Poor numbers: how we are misled by African development statistics and what to do about it*. Ithaca and London: Cornell University Press.
- Karlan D. (2015), *The Multi-Lateral Development Banks: A Strategy for Generating Increased Return on Investment*, Washington D.C.: US Congress, Testimony before the US House Committee on Financial Services, online at [www.poverty-action.org/sites/default/files/Dean%20Karlan\\_Testimony.pdf](http://www.poverty-action.org/sites/default/files/Dean%20Karlan_Testimony.pdf)
- Karlan D.S. and J. Appel (2012), *More than Good Intentions; Improving the Ways the World's Poor Borrow, Save, Farm, Learn, and Stay Healthy*, New York: Dutton Press.
- Kosters M. and J. Van der Heijden (2015), “From mechanism to virtue: Evaluating Nudge theory” *Evaluation* 21(3): 276-291.
- Labrousse A. (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): 2-32.
- Latour B. (1999), *Pandora's Hope: Essays on the Reality of Science Studies*, Cambridge: Harvard University Press.
- Levine R, R. Polimeni and I. Ramage (2016), “Insuring health or insuring wealth? An experimental evaluation of health insurance in rural Cambodia” *Journal of Development Economics*, 119: 1-15.
- List J.A. and R Metcalfe (2014), “Field experiments in the developed world: an introduction” *Oxford Review of Economic Policy* 30: 585-596.
- McKenzie D. (2016), “Have RCTs taken over development economics?” *World Bank Blog on Impact Evaluations*, online at: <http://blogs.worldbank.org/impactevaluations/have-rcts-taken-over-development-economics>.
- Miguel E. and M. Kremer (2004), “Worms: identifying impacts on education and health in the presence of treatment externalities” *Econometrica* 72 (1): 159-217.

- Milonakis D. and B. Fine (2009), *From political economy to economics*, London: Routledge.
- Miranda J., S. Sabet S. and A.N. Brown (2016), “Is impact evaluation still on the rise?” *Evidence Matters: Improving development and practice*, online at: <http://blogs.3ieimpact.org/is-impact-evaluation-still-on-the-rise/>.
- Moffitt R. A. (2004), “The Role of Randomized Field Trials in Social Science Research A Perspective from Evaluations of Reforms of Social Welfare Programs” *American Behavioral Scientist* 47 (5): 506-540.
- Moyo D. (2009), *L'aide fatale, Les ravages d'une aide inutile et de nouvelles solutions pour l'Afrique*, Paris: Hachette.
- Morvant-Roux S., I. Guérin, M. Roesch and J.Y. Moisseron (2014), “Adding value to randomization with qualitative analysis: the case of microcredit in rural Morocco” *World Development* 56: 302-312.
- Muralidharan K. and V. Sundararaman (2011), “Teacher Performance Pay: Experimental Evidence from India” *Journal of Political Economy* 119 (1): 39-77.
- Naudet J.-D. (2006), “Les OMD et l'aide de cinquième génération” *Afrique contemporaine* 218 (2): 141-174.
- Nubukpo K. (2012), “Esther Duflo, ou l'économie expliquée aux pauvres d'esprit”, *Blog Alternatives Economiques : L'actualité vue par Kako Kubukpo*.
- Oakley A. (2000), “A Historical Perspective on the Use of Randomized Trials in Social Science Settings” *Crime & Delinquency* 46 (3): 315-329.
- O'Laughlin B. (2015), “Trapped in the prison of the proximate: structural HIV/AIDS prevention in southern Africa” *Review of African Political Economy* 42 (145): 342-361.
- Picciotto R. (2012), “When will the bubble burst?” *Evaluation* 18 (2): 213-229.
- Pritchett L. (2012), “Impact evaluation and political economy: what does the ‘conditional’ in ‘conditional cash transfers’ accomplish?”, Center for Global Development blog, online at: <https://www.cgdev.org/blog/impact-evaluation-and-political-economy-what-does-%E2%80%9Cconditional%E2%80%9D-%E2%80%9Cconditional-cash-transfers%E2%80%9D>.
- Pritchett L., S. Samji and J. Hammer (2013) “It’s All About MeE: Using Structured Experiential Learning (“e”) to Crawl the Design Space”, *Harvard Kennedy School*, RWP 13-012.
- Pritchett L., Sandefur J. (2015), “Learning from Experiments When Context Matters” *American Economic Review* 105(5): 471-475.
- Pritchett L. and J. Sandefur (2013), “Context Matters for Size: Why External Validity Claims and Development Practice Don’t Mix”, *Center for Global Development*, Working Paper.
- Quentin A. and I. Guérin (2013), « La randomisation à l'épreuve du terrain. L'expérience de la micro-assurance au Cambodge" *Revue Tiers Monde* 213: 179-200.
- Ravallion M. (2009), “Evaluation in the practice of development” *The World Bank Research Observer* 24: 29–53.

- Rodrik D. (2008), "The new development economics: we shall experiment, but how shall we learn?", John F. Kennedy School of Government, Harvard University.
- Rosholm M. and L. Skipper (2009), "Is labour market training a curse for the unemployed? Evidence from a social experiment" *Journal of Applied Economics* 24: 338–365.
- Scriven M. (2008), "A Summative Evaluation of RCT Methodology: An Alternative Approach to Causal Research" *Journal of MultiDisciplinary Evaluation* 5(9): 11-24.
- Savedoff W. D., R. Levine and N. Birdsall (Ed.) (2006), *When Will We Ever Learn? Improving Lives Through Impact Evaluation*, Washington D.C.: Center for Global Development.
- Schulz, K. F, I. Chalmers, R. J Hayes, and D. G Altman (1995), "Empirical Evidence of Bias" *JAMA: The Journal of the American Medical Association* 273(5): 408-412.
- Shaffer P. (2011), "Against Excessive Rhetoric in Impact Assessment: Overstating the Case for Randomised Controlled Experiments" *Journal of Development Studies* 47(11): 1619–1635.
- Shah N. B., P. Wang, A. Fraker and D. Gastfriend (2015), "Evaluations with impact: decision-focused impact evaluation as a practical policy making tool", *3IE*, Working Paper 25.
- Sharma M. and M. Zeller (1997), "Repayment performance in group-based credit programs in Bangladesh: An empirical analysis", *World Development*, 25(10), pp. 1731-1742.
- Teele D. L. (2014), *Field experiments and their critics. Essays on the Uses and Abuses of Experimentation in the Social Sciences*, New Haven & London: Yale University Press.
- Vivalt E. (2016), *How Much Can We Generalize from Impact Evaluations?*, Stanford: Stanford University.
- WHO (2014), *Preventing diarrhoea through better water, sanitation and hygiene: exposures and impacts in low- and middle-income countries*, Geneva: World Health Organization.
- Woolcock M. (2013), "Using case studies to explore the external validity of 'complex' development interventions" *Evaluation* 19(3): 229-248.
- Wolf, J., A. Prüss Ustün, O. Cumming, J. Bartram, S. Bonjour, S. Cairncross, T. Clasen, J. M Colford, V. Curtis and J. France (2014), "Systematic review: Assessing the impact of drinking water and sanitation on diarrhoeal disease in low and middle income settings: systematic review and meta regression" *Tropical Medicine & International Health* 19: 928-942.
- Young A. (2015), "Channelling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results", *LSE*, Working Paper.

## Dernières parutions de la collection *Papiers de Recherche de l'AFD*

---

Les *Papiers de Recherche* de l'AFD sont disponibles sur : <http://librairie.afd.fr/>

- # 2017-33 DAVID, A. and M.-A. MAROUANI (2017), "Migration patterns and labor market outcomes in Tunisia", AFD Research Paper Series, n° 2017-33, February.
- # 2017-34 D'AIGLEPIERRE, R. and A. BAUER (2017), "The choice of Arabo-Islamic education in sub-Saharan Africa: findings from a comparative study", AFD Research Paper Series, n° 2017-34, March.
- # 2017-35 AVITABLEY, C., BOBBA, M. and M. PARIGUANAX (2017), "High School Track Choice and Liquidity Constraints: Evidence from Urban Mexico", AFD Research Paper Series, n° 2017-35, March.
- # 2017-36 BOBBA, M. and V. FRISANCHO (2017), "Learning about Oneself: The Effects of Performance Feedback on School Choice", AFD Research Paper Series, n° 2017-36, March.
- # 2017-37 WAGNER, L. and R. D'AIGLEPIERRE (2017), "Macroeconomic Crisis, Primary Education and Aid Effectiveness", AFD Research Paper Series, n° 2017-37, February.
- # 2017-38 BAUER, A. (2017), "Redistributive Programs' Implementation: Do Political Incentives Matter?", AFD Research Paper Series, n° 2017-38, March.
- # 2017-39 D'AIGLEPIERRE, R. and A. BAUER (2017), "Explaining the Development of Private Education: the Effect of Public Expenditure on Education", AFD Research Paper Series, n° 2017-39, March.
- # 2017-40 NOCUS, I., GUIMARD, P. and A. FLORIN (2017), "Evaluation of the « *Ecole et langues nationales en Afrique* » program: Methodological Aspects and Interim Assessment", AFD Research Paper Series, n° 2017-40, February.
- # 2017-41 ALTET, M., N. SALL, M. SECK and I. OUEDRAOGO (2017), "Teacher training based on the results of research on their actual practices in the context of Sub-Saharan Africa: The OPERA project in Burkina Faso, issues and challenges", AFD Research Paper Series, n° 2017-41, February.
- # 2017-42 GIRAUD, G. and M. GRASSELLI (2017), "The macrodynamics of household debt, growth, and inequality", AFD Research Paper Series, n° 2017-42, April.
- # 2016-43 FEUBI PAMEN, E.-P. and M. KUEPIE (2017), "An Application of the Alkire-Foster's Multidimensional Poverty Index to Data from Madagascar: Taking Into Account the Dimensions of Employment and Gender Inequality", AFD Research Paper Series, No. 2017-43, April.