

Les expérimentations aléatoires en microfinance

Miracle ou mirage ?

Florent BÉDÉCARRATS, Isabelle GUÉRIN et François ROUBAUD

Introduction

L'essor du microcrédit et la généralisation des évaluations par assignation aléatoire (*Randomized Controlled Trials* – RCT) ont marqué deux avancées majeures des politiques de développement de réduction de la pauvreté au cours des dernières décennies (CLING *et al.*, 2003). L'essor du microcrédit a eu lieu dans les années 1990. Son point culminant date du début des années 2000, avec le lancement par l'ONU de l'année internationale du microcrédit (2005) et l'attribution du prix Nobel de la paix à Mohammad Yunus et à la Grameen Bank, qu'il a fondée. Les RCT ont connu un succès retentissant dix ans plus tard, lorsqu'elles sont devenues la méthode de référence par excellence pour les évaluations d'impact et quand, en 2019, un autre prix Nobel (d'économie cette fois-ci) a été décerné à Esther Duflo, Abijit Banerjee et Michael Kremer, chefs de file du mouvement RCT. De fait, ces deux avancées sont étroitement corrélées : le microcrédit a été l'un des thèmes phares, un sujet emblématique, des interventions évaluées par les expérimentations randomisées dans le domaine du développement.

Ce chapitre présente un examen détaillé des RCT sur le microcrédit en s'appuyant sur un large éventail d'outils analytiques employés en statistique, en économie politique, en sociologie et en anthropologie du développement. Il se concentre principalement sur le numéro spécial (ci-après dénommé « numéro spécial ») publié en 2015 dans une grande revue d'économie, *American Economic Journal: Applied Economics* (*AEJ:AE*). Ce numéro spécial réunit six RCT sur le microcrédit et les articles sont préfacés par une introduction générale (ci-après appelée « introduction générale ») qui tire des conclusions communes. Le numéro spécial

a rencontré un large écho dans les cercles académiques et professionnels, tant et si bien qu'il a tendance à être considéré comme la conclusion définitive sur les impacts (limités) du microcrédit. Mais l'est-il vraiment ?

Nous abordons ce numéro spécial sous deux angles : (1) *top-down*, avec un test sur un cas précis (microcrédit) illustrant les critiques générales faites contre les RCT, en particulier celles formulées par les auteurs dans un précédent article (BÉDÉCARRATS *et al.*, 2019b) et (2) *bottom-up*, avec une étude de la mise en œuvre des RCT sur le terrain. Nous prenons comme point de départ une réplique de l'une des six RCT abordées dans le numéro spécial : la RCT conduite dans le Maroc rural (BÉDÉCARRATS *et al.*, 2019a ; 2019b), qui joue un rôle central dans l'« économie » du numéro spécial. Nous élargissons ensuite la perspective du cas marocain pour adopter un angle plus général en identifiant les invariants présents dans d'autres RCT et en déterminant les particularités de chacune d'entre elles. De façon plus générale, la principale question que nous posons dans ce chapitre est la suivante : « Quels enseignements peut-on tirer des RCT sur le microcrédit et comment peut-on expliquer leur succès mondial alors qu'elles souffrent d'un manque évident de rigueur ? »

Le reste de ce chapitre est organisé comme suit. Après avoir résumé les principales caractéristiques des six expérimentations, la deuxième partie présente leurs principaux résultats et replace le numéro spécial dans le contexte général du poids et du rôle du microcrédit dans l'industrie des RCT. La troisième partie adopte un point de vue comparatif pour identifier les principales critiques techniques qui peuvent être formulées contre ce corpus d'expérimentations, tant en termes de validité interne que de validité externe, ainsi que les problèmes éthiques soulevés. Au-delà de la méthode et des résultats quantitatifs, la quatrième partie analyse les interprétations proposées par les auteurs (notamment dans l'introduction générale) et leur théorie sous-jacente du changement. Dans la conclusion, nous proposons une interprétation du hiatus évoqué ci-dessus – un succès de grande envergure malgré des lacunes importantes – et tirons des enseignements plus généraux de notre travail.

RCT et microcrédit : un produit phare en désuétude ?

Le microcrédit figure parmi les principaux services fournis par la microfinance, l'un des secteurs les plus fréquemment évalués par les RCT. Cette importance est illustrée par la base de données en ligne des RCT gérées par le Abdul Latif Jameel Poverty Action Lab (J-PAL) – un centre de recherche mondial qui préconise cette méthode pour la réduction de la pauvreté et qui, de fait, est le principal fournisseur et promoteur des RCT. En 2010, cette base de données comptait 233 RCT, dont 32 % classées sous le label « microfinance » (BÉDÉCARRATS,

2012). Depuis lors, le J-PAL a revu son système de classement en élargissant ses catégories et compte actuellement 287 RCT classées en « finance » sur un total de 978 RCT¹. De fait, la finance est le secteur d'intérêt par excellence du J-PAL, devant l'éducation (233 RCT) et « l'économie politique et la gouvernance » (216 RCT). Bien que la microfinance ne soit qu'un sous-ensemble des RCT « finance », le J-PAL produit un grand nombre d'évaluations d'impact sur le sujet. Les RCT sur la microfinance et l'industrie générale des RCT ont véritablement pris leur envol au milieu des années 2000 (BÉDÉCARRATS *et al.*, 2019b ; Ravallion, chap. 1, ce volume). Depuis, le nombre de RCT consacrées à la microfinance a fortement régressé, alors que les RCT en général ont poursuivi leur expansion (fig. 1). Notons que le comptage des RCT conduites à travers le monde est un défi. Nos estimations sont fondées sur la base de données d'évaluations d'impact en ligne de l'International Initiative for Impact Evaluation (3ie), complétée par BÉDÉCARRATS (2012) et sur la base de données d'évaluation en ligne du J-PAL². Comme l'indique la fig. 1a, l'impact de la microfinance constitue de longue date un sujet controversé, source de nombreuses évaluations d'impact non expérimentales. Alors que les méthodes expérimentales produisent des preuves empiriques quantitatives théoriquement plus solides, les études non expérimentales offrent pléthore de preuves pertinentes. Une forte hausse des évaluations expérimentales a également été observée, coïncidant avec une diminution nette des évaluations non expérimentales, même si ces tendances peuvent être marginalement exagérées par l'omission des études les plus récentes dans les référentiels que nous avons utilisés. La fig. 1b montre par ailleurs que la microfinance a constitué un thème prédominant du mouvement *randomista*³ jusqu'en 2013, mais que cet intérêt s'est depuis affaibli. Le déclin observé dans la seconde moitié des années 2010, après le pic de la première moitié de cette décennie est quelque peu intrigant : est-il dû à une inversion de tendance ou est-ce parce qu'il n'y a plus grand-chose à dire sur ce sujet, désormais trop rebattu ? C'est un point que nous traiterons plus en détail par la suite.

1. Source : site internet de The Abdul Lateef Jameel Poverty Action Lab : www.povertyactionlab.org/evaluations.

2. La base de données d'évaluations d'impact en ligne du 3ie constitue le principal catalogue de résultats d'évaluations d'impact conduites sur des interventions de développement (<https://www.3ieimpact.org/evidence-hub/impact-evaluation-repository>, consulté le 13/10/2019). Le 3ie a tendance à omettre des évaluations non expérimentales et son travail d'inventaire semble avoir perdu de sa vigueur ces dernières années, les références se faisant plus rares à partir de 2015. Nous avons complété les données du 3ie par les évaluations d'impact recensées dans BÉDÉCARRATS (2012) et par les références incluses dans la base de données d'évaluation du J-PAL. Les références ont été appariées afin d'éviter de compter deux fois les mêmes évaluations. La fig. 1b est basée sur les références recensées dans la base de données d'évaluation en ligne du J-PAL (<https://www.povertyactionlab.org/evaluations>, consulté 18/10/2019). Le terme « finance » dans la clé est le label assigné par le J-PAL à l'évaluation enregistrée. Les auteurs ont attribué le label « microfinance » après examen des résumés de toutes les évaluations enregistrées en « finance » sur le site du J-PAL. Les dates de la fig. 1b indiquent l'année d'achèvement de l'expérimentation, tandis que les dates de fig. 1a indiquent l'année de publication de ses résultats.

3. Nous appelons « *randomistas* » les promoteurs de RCT qui sont convaincus que les RCT constituent le seul moyen d'évaluer rigoureusement l'impact dans l'évaluation et sont en tout point supérieures aux autres méthodologies.

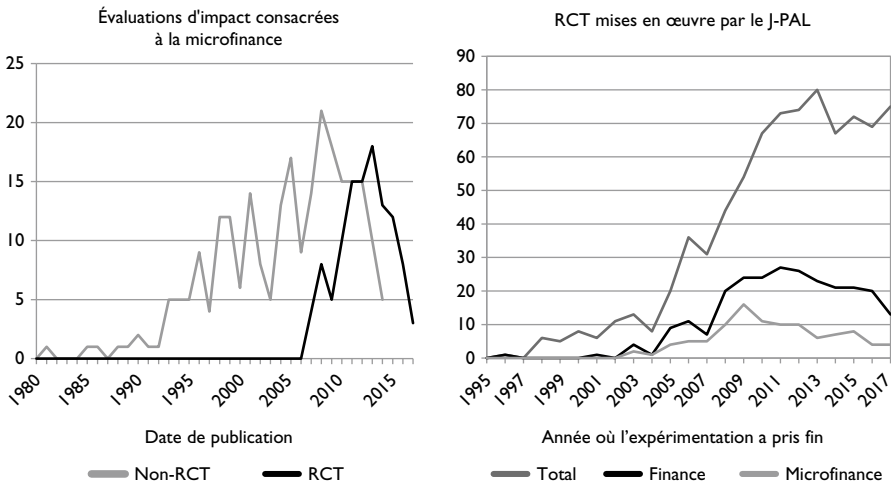


Figure 1
RCT sur la microfinance.

Source : Florent Bédécarrats, Isabelle Guérin et François Roubaud, sur la base du référentiel d'évaluations 3ie (2019), du référentiel d'évaluations J-PAL et de BÉDÉCARRATS (2012) pour le panel Ia ; et sur la base du référentiel d'évaluations en ligne J-PAL pour le panel Ib.

C'est à l'apogée des RCT consacrées à la microfinance qu'un numéro spécial a été publié en 2015 dans l'*American Economic Journal: Applied Economics*, présentant six RCT sur le microcrédit (BANERJEE *et al.*, 2015c). Pour les chefs de file du mouvement RCT, ce numéro spécial est jugé comme une contribution décisive, permettant de clore un débat (OGDEN, 2017) qui a longtemps fait rage tant dans les cercles académiques que parmi les bailleurs et les décideurs politiques. Ce numéro spécial a rapidement bénéficié d'une couverture massive, comme en témoignent les 3 607 citations de ses articles parus dans d'autres publications scientifiques⁴. Soucieux d'en tirer des leçons opérationnelles à l'intention des décideurs politiques, le J-PAL et l'Innovations for Poverty Action (IPA) en ont tiré une note de synthèse suggérant des conclusions générales et universelles, valables pour le microcrédit dans le monde entier (J-PAL et IPA, 2015). Certains chercheurs ont même spéculé que ce pourrait être là le « dernier mot sur le microcrédit » (SANDEFUR, 2015).

Si l'on examine de plus près l'impact académique du numéro spécial de l'*AEJ:AE*, le résultat est saisissant. Selon Google Scholar (consulté le 13/10/2019), l'introduction générale a été citée pas moins de 527 fois. C'est là une belle performance, bien que loin derrière l'article de BANERJEE *et al.* (2015b) sur le programme de microcrédit Spandana en Inde (1 813 citations). Les cinq autres articles ont également obtenu de très bons scores : 320 citations pour ANGELUCCI, KARLAN et ZINMAN (2015) sur le Compartamos Banco au Mexique, 298 pour CRÉPON *et al.* (2015) sur *Al Amana* dans le Maroc rural, 225 pour ATTANASIO *et al.* (2015) sur

4. https://scholar.google.fr/scholar?as_q=microcredit+OR+microfinance&as_epq=&as_oq=&as_eq=&as_occt=any&as_sauthors=&as_publication=american+economic+journal+applied&as_ylo=2015&as_yhi=2015&hl=en&as_sdt=0%2C5.

la Mongolie, 214 pour AUGSBURG *et al.* (2015) sur la Bosnie-Herzégovine et 210 pour TAROZZI *et al.* (2015) sur l'Éthiopie. À titre de comparaison, l'article de PITT et KHANDKER (1998), cité par ROODMAN et MORDUCH (2014) comme étant l'article empirique sur le microcrédit le plus cité de tous les temps, affiche 1 956 citations, plus de vingt ans après sa publication.

Outre ces citations directes, l'impact du numéro spécial s'illustre par des citations en cascade (à l'instar de tout article), mais aussi par des examens systématiques ou des méta-analyses qui, pour la plupart, trouvent dans le numéro spécial l'essentiel de leur corpus de preuves (BRODY *et al.*, 2015 ; BUERA *et al.*, 2015 ; DEMIRGUC-KUNT *et al.*, 2017 ; CHERNOZHUKOV *et al.*, 2018 ; MEAGER, 2019). Il convient de mentionner en particulier l'article publié dans la prestigieuse revue *Science* en 2015 (BANERJEE *et al.*, 2015a, cité 484 fois)⁵. Cet article traite en détail du numéro spécial, en mettant en exergue les mérites comparatifs d'une approche distincte (programmes dits de « graduation »)⁶.

Enfin, les résultats du numéro spécial ont largement dépassé les cercles académiques pour pénétrer le monde des praticiens de la microfinance (J-PAL et IPA, 2015). Le Consultative Group to Assist the Poor (CGAP), qui joue un rôle de premier plan dans la diffusion des bonnes pratiques dans le secteur du microcrédit, l'a commenté avant même sa publication (CULL *et al.*, 2014). Pour de nombreux praticiens (que l'un d'entre nous rencontre régulièrement dans des conférences et sur le terrain), les résultats du numéro spécial relèvent désormais du sens commun.

En définitive, qu'elles soient jugées sur le nombre de RCT ou sur la diffusion de leurs résultats, les évaluations d'impact dans le domaine de la microfinance et, en particulier, du microcrédit, apparaissent comme le produit phare du mouvement *randomista*, et le numéro spécial comme le prototype par excellence de ce produit phare.

Gros plan sur la conception du numéro spécial de l'AEJ:AE

Le numéro spécial comporte six articles sur six RCT sur le microcrédit, menées par six équipes affiliées au J-PAL dans six pays différents (Bosnie-Herzégovine, Éthiopie, Inde, Mexique, Mongolie et Maroc) à peu près au même moment (de 2006 à 2012). Ces articles sont précédés d'une introduction générale qui tire des enseignements généraux de cette expérience collective. Le numéro spécial puise sa force dans un processus d'harmonisation organisé en aval par le journal en vue de sa publication⁷. Un plan d'analyse commun a été mis au point pour faciliter les comparaisons. Dans la mesure du possible, l'impact du microcrédit a été estimé

5. Ce n'est pas la première fois que *Science* ouvre ses colonnes aux RCT sur le microcrédit (KARLAN et ZINMAN, 2011).

6. Les programmes dits de « graduation » consistent en un ensemble séquencé d'interventions ciblant les ultra-pauvres et extrêmement pauvres qui sont enfermés dans les situations de pauvreté les plus difficiles.

7. « Grâce aux efforts déployés par les six équipes de recherche et la rédactrice en chef, Esther Duflo, pour rendre les articles facilement comparables, il a été relativement aisé de tirer des enseignements des six études » (BANERJEE *et al.*, 2015c : 2).

selon la même méthodologie économétrique afin d'obtenir un ensemble de résultats communs, eux-mêmes tous calculés de la même manière. C'était la première fois qu'un tel effort de mutualisation était déployé à cette échelle. Dans l'optique d'une montée en généralisation, cette mutualisation constitue un avantage décisif.

Ainsi, le numéro spécial n'est pas seulement déterminant en termes de résultats, il marque aussi un changement de « bonnes pratiques » de la part des promoteurs des RCT. Le numéro cherche en effet à remédier à un certain nombre de limites. Pour la première fois, le numéro dans son ensemble, et l'introduction générale en particulier, apporte des éléments de réponse à cinq types de critiques récurrentes du mouvement pro-RCT (BÉDÉCARRATS *et al.*, 2019b) : un modèle théorique est développé en réponse à la critique accusant les RCT d'empirisme agnostique ; une analyse coût-bénéfice est proposée pour répondre à la question de l'efficacité, afin de dépasser le simple impact causal ; les questions du taux d'adhésion, de la précision des estimateurs et de l'hétérogénéité des traitements sont reconnues et discutées ; la diversité contextuelle est abordée par une série de paramètres, de produits et d'institutions couverts par les six articles, permettant aux éditeurs du numéro spécial d'affirmer que leur échantillon est « assez représentatif de l'industrie/du mouvement du microcrédit dans le monde » (BANERJEE *et al.*, 2015c : 2) ; et, enfin, le numéro spécial se propose de mettre à disposition les bases de données originales en réponse aux requêtes de réplcation et afin de faciliter les méta-analyses.

Décrivons brièvement les six RCT. En dépit d'un processus d'harmonisation en amont (traitement et analyse des données), les expériences diffèrent grandement dans leurs protocoles. Les types de produits de microcrédit, les institutions de microfinance (IMF), l'unité des procédures de randomisation, etc. varient d'une RCT à l'autre. Les auteurs interprètent cette diversité en partant de l'hypothèse que la similitude des résultats dans ce large éventail d'environnements est une garantie de leur robustesse et met donc en évidence les propriétés génériques des impacts du microcrédit. En d'autres mots, ils entendent répondre à la critique récurrente déplorant le manque de validité externe des RCT.

L'introduction générale présente en détail les principales caractéristiques des six RCT, résumées dans le tabl. 1. Les IMF varient en taille, certaines ont une vocation commerciale et d'autres non. Toutes sortes de produits sont proposés : des prêts collectifs et individuels, des remboursements hebdomadaires et mensuels, un taux d'intérêt annuel allant de 12 à 110 % (en moyenne) et un montant de prêt (moyen) représentant 6 à 118 % du revenu mensuel. La moitié des programmes de microcrédit ciblent des femmes. Au niveau géographique, une RCT est exclusivement urbaine (Inde), trois sont exclusivement rurales (Éthiopie, Mongolie et Maroc) et les deux dernières couvrent les deux types de zones. Il convient de noter que, dans tous les cas, les critères d'éligibilité des clients sont *ad hoc* : ils dépendent à la fois des règles internes de chaque IMF et des paramètres de chaque RCT. Ce faisant, les populations cibles sont extrêmement spécifiques (voire uniques), ce qui restreint les possibilités d'inférence et d'extrapolation à des populations plus larges ; nous reviendrons sur ce point dans la troisième partie.

Tableau 1
Principales caractéristiques des six RCT.

	Bosnie- Herzégovine	Éthiopie	Inde	Mexique	Mongolie	Maroc
Taux annuel effectif global (TAEG)	22 %	12 %	24 %	110 %	27 %	14 %
Garantie	Individuelle	Solidaire	Solidaire	Solidaire	Les deux	Solidaire
Prêt/revenu du ménage moyen	9 %	118 %	22 %	6 %	43 %	21 %
Sexe des clients potentiels	Les deux	Les deux	Féminin	Féminin	Féminin	Les deux
Éligibilité au prêt (entre autres)	Garantie solide, capacité de remboursement, solvabilité...	Statut de pauvreté, <i>business plan</i> ...	Entre 18 et 59 ans, justificatif de propriété du logement...	Entre 18 et 60 ans, carte d'identité en cours de validité, justificatif de domicile...	Actifs < 869 \$ Bénéfices < 174 \$/mois	Entre 18 et 70 ans, carte d'identité, activité agricole autre qu'élevage...
Couverture géographique (urbaine/rurale)	Les deux	Rurale	Urbaine	Les deux	Rurale	Rurale
Couverture géographique (en nombre de régions ou villes)	14 (à l'échelle nationale)	2 (Ouest)	1 (ville)	4 (NC Sonora)	5 (Nord)	11 (à l'échelle nationale)
Unité de randomisation	Individuelle	Association	Quartier	Quartier et village	Village	Village
Unité d'échantillonnage finale	Candidat risqué et jugé insolvable...	Ménages aléatoires	Ménage avec ≥ 1 femme ≥ 3 ans dans la zone...	Avoir une entreprise ou souhaiter en avoir une...	Être intéressé par l'obtention d'un prêt...	Ménage considéré comme emprunteur potentiel...
Taille de l'échantillon (enquête finale)	995	6 263	6 862	16 560	964	5 551

Source : Florent Bédécarrats, Isabelle Guérin et François Roubaud, sur la base de BANERJEE *et al.* (2015c, tabl. 1 et 2).

Exploration détaillée du numéro spécial de l'*AEJ:AE* : principaux résultats

L'introduction générale tire sept grands enseignements de l'exercice. En premier lieu, le faible niveau d'adhésion est une constante dans toutes les études, sauf en Bosnie-Herzégovine, ce qui amène à la conclusion que le microcrédit ne saurait faire figure de panacée pour sortir les populations de la pauvreté. Une conséquence préjudiciable de cette faible adhésion est qu'elle pose un problème de puissance statistique et remet en cause la stratégie d'identification des RCT. Toutefois, l'introduction générale prétend que les RCT du Maroc, de l'Inde et du Mexique apportent de nouveaux éléments pour remédier à ces lacunes (prévision du taux d'adhésion et stratégie d'échantillonnage). Deuxièmement, et dans le prolongement du point précédent, il est particulièrement difficile de prévoir le taux d'adhésion, et aucune étude n'est parvenue à le faire de façon satisfaisante. Troisièmement, et c'est probablement la principale conclusion, l'accès au microcrédit n'est pas transformateur, que ce soit au niveau des performances des micro-entreprises ou des conditions de vie des ménages – y compris concernant le « bien-être social » et « l'autonomisation » des femmes –, du moins en moyenne. Le seul résultat robuste pour la consommation est une baisse des dépenses dites « facultatives », à savoir, selon les auteurs, les « biens de tentation, loisirs/divertissements/célébrations » (BANERJEE *et al.*, 2015c : 13). Quatrièmement, seul l'investissement des entreprises est augmenté par le microcrédit, ce qui montre que ce dernier stimule les intentions des micro-entrepreneurs visant à développer leur entreprise. Cinquièmement, d'autres effets modestes, mais potentiellement importants sont relevés : la liberté de choix en particulier. Sixièmement, bien que le microcrédit ne soit pas transformateur, il n'a pas non plus d'effets catastrophiques, ce qui place les partisans et les adversaires du microcrédit sur un pied d'égalité. Enfin, le septième enseignement concerne la présomption d'hétérogénéité de l'impact du microcrédit, qui pourrait être positif (voire transformateur) pour certains (la couche supérieure) et négatif pour d'autres. Cela nous ramène à la question de la puissance statistique, de la taille d'échantillon requise pour estimer correctement les impacts, et de la représentativité des populations ciblées. Le tabl. 2, basé sur l'introduction générale et le J-PAL et IPA (2015), résume les résultats obtenus par les six RCT au regard des principaux effets suivis.

En conclusion, le numéro spécial est considéré par beaucoup, à commencer par les auteurs eux-mêmes (OGDEN, 2017), comme la synthèse la plus complète qui soit sur l'impact du microcrédit. Ses conclusions générales n'ont guère été remises en cause depuis sa publication en 2015 (pour les exceptions, voir WYDICK, 2016 ; DAHAL et FIALA, 2020). D'une certaine manière, il fige l'état des connaissances acquises sur les impacts causaux du microcrédit et sur son rôle dans le développement et l'éradication de la pauvreté. Pour les éditeurs de l'*AEJ:AE*, et les articles ultérieurs traitant des six RCT, le numéro spécial va encore bien plus loin. Il est salué pour avoir repoussé les frontières de la connaissance scientifique, tant en ce qui concerne le microcrédit que la méthode RCT. Trois papiers, postérieurs au numéro spécial et traitant directement du même ensemble de RCT, illustrent bien cet état de fait. L'article de MEAGER (2019), publié lui aussi dans l'*AEJ:AE*, confirme que le numéro spécial reste

incontournable. Cet article reprend les six RCT du numéro spécial (plus une RCT aux Philippines ; KARLAN et ZINMAN, 2011) pour réestimer l'impact général sur les principales variables et répondre à la question de la validité externe à l'appui d'une méthode innovante (une analyse hiérarchique bayésienne). Il y a ensuite CHERNOZHUKOV *et al.* (2018), qui applique une méthode de double apprentissage automatique pour étudier l'hétérogénéité de cet ensemble de données. Un troisième exemple est celui de BANERJEE *et al.* (2019a), publié au moment où ce chapitre était en cours de rédaction. Leur article s'appuie sur une troisième vague d'enquête pour la RCT indienne de Spandana. Tout en répondant à certaines des critiques formulées contre les RCT (en abordant le traitement hétérogène, en allongeant la durée et en développant un modèle théorique), l'article puise largement dans le numéro spécial, qui est ici présenté comme la somme des connaissances réunies à ce jour sur le microcrédit. Cet article n'est certainement pas le dernier. Dans la même veine, CRÉPON *et al.* (2015) annoncent dans la conclusion de leur article une troisième vague d'enquêtes pour la RCT marocaine afin d'évaluer l'impact à long terme du microcrédit⁸.

Tableau 2
Principaux résultats des six RCT.

	Bosnie-Herzégovine	Éthiopie	Inde	Mexique	Mongolie	Maroc
Propriété d'entreprise	Positif	n.s.	n.s.	n. s.	Positif	n. s.
Revenus d'entreprise	n.s.	n.s.	n. s.	Positif	n. s.	Positif
Actifs d'entreprise	Positif	–	Positif	–	Positif	Positif
Investissements d'entreprise	n.s.	n. s.	Positif	Positif	–	Positif
Bénéfices d'entreprise	–	–	–	–	–	Positif
Revenus des ménages	n.s.	n.s.	n.s.	n.s.	n.s.	n. s.
Consommation des ménages	n. s.	Négatif	–	Négatif	Positif	–
Consommation des ménages en biens de tentation	Négatif	–	Négatif	Négatif	n. s.	Négatif
Bien-être social	n.s.	n.s.	n. s.	Positif	–	n. s.
Autonomisation des femmes	–	n. s.	–	Positif	–	–

Source : Florent Bédécarrats, Isabelle Guérin et François Roubaud, sur la base de J-PAL et IPA (2015) ; BANERJEE *et al.*, (2015c).

Note : n.s. : non significatif à 10 % ; « – » : pas de données.

8. « Nous procédons actuellement à un suivi auprès des ménages, maintenant qu'un laps de temps beaucoup plus long s'est écoulé, pour vérifier si l'investissement dans les actifs commerciaux s'est révélé rentable sur le long terme » (CRÉPON *et al.*, 2015 : 148).

Validité et portée du numéro spécial : évaluation critique

Dans la littérature, les RCT sont évaluées sous deux angles principaux : la validité externe et la validité interne. La validité externe est essentielle lorsqu'il s'agit de transposer, d'informer et de concevoir des politiques publiques à une échelle plus large (nationale ou régionale) et de tester une théorie. La validité interne est généralement prise pour un fait acquis avec les RCT, et considérée comme leur principal point fort par rapport aux autres méthodes. Si cette caractéristique peut être vraie en théorie, les contraintes de mise en œuvre sur le terrain peuvent remettre en cause ces conditions idéales, un aspect jusqu'ici passé sous silence.

Validité interne

L'évaluation de la validité interne des RCT requiert des investigations sur l'élaboration, et le « bricolage », des RCT sur le terrain. Nous nous sommes livrés à cet exercice exigeant sur l'étude marocaine (CRÉPON *et al.*, 2015). Nous présentons ci-dessous les principaux résultats des deux articles complémentaires que nous avons produits à partir de cette étude (BÉDÉCARRATS *et al.*, 2019a ; 2019b).

Le cas emblématique de la RCT marocaine

De 2006 à 2010, une équipe de recherche du J-PAL a conduit une RCT dans le Maroc rural pour mesurer l'impact du microcrédit fourni par *Al Amana*, qui était à l'époque la principale IMF sur un marché marocain en pleine phase d'expansion.

Nous avons répliqué l'enquête de Crépon *et al.* et avons identifié un certain nombre de problèmes qui remettent en cause leurs conclusions (BÉDÉCARRATS *et al.*, 2019a). Les chercheurs ont utilisé des procédures et des seuils incohérents pour le *trimming* et leurs résultats sont lourdement tributaires de la manière dont les valeurs extrêmes ont été exclues. CRÉPON *et al.* (2015) font état d'un échantillon équilibré dans l'enquête initiale (EI), après avoir éliminé les valeurs extrêmes de 24 variables sur 459 observations (10,3 % de l'échantillon). Mais, dans l'enquête finale (EF), ils ont éliminé différemment 27 observations (0,5 % de l'échantillon) en les supprimant complètement. Le fait d'abaisser ce seuil dans l'enquête finale de seulement 0,2 % (en éliminant une douzaine d'observations en plus ou en moins) produit des résultats radicalement différents en termes de ventes, de dépenses, d'investissements et de bénéficiaires. Aucun autre seuil appliqué pour le *trimming* n'aurait donné des résultats conformes aux conclusions publiées et aucun autre article du même numéro spécial n'a eu recours à une méthode ou à un seuil similaire.

Nous avons constaté des déséquilibres substantiels et significatifs dans l'enquête initiale pour un certain nombre de variables importantes, dont les variables de résultats de la RCT. Dans ce contexte, nous avons estimé des « effets de traitement » peu plausibles sur certaines variables, par exemple le chef de ménage, le

sexe et la langue parlée. Nous avons en outre documenté de nombreuses erreurs de codage. Par exemple, l'évaluation des actifs agricoles dans l'enquête finale a omis deux types d'actifs (les tracteurs et les moissonneuses), qui se trouvent être les actifs les plus chers des ménages interrogés. Or, l'inclusion de tracteurs et de moissonneuses dans l'évaluation des actifs augmente la valeur moyenne des actifs agricoles par ménage de 470 % (de 1 377 à 5 111 dirhams). Les erreurs de codage identifiées ont altéré environ 80 % des observations.

Les incohérences des mesures du crédit méritent une attention particulière, car elles revêtent une importance primordiale pour caractériser le traitement évalué par cette expérimentation. CRÉPON *et al.* (2015) comparent les données administratives (celles de l'IMF) aux données de l'enquête, en préférant se baser sur le taux d'adhésion de 17 % des données administratives, contre les 11 % de l'enquête. Ils affirment que la population marocaine sous-estime les emprunts en raison de la honte religieuse qu'ils impliquent. Cependant, cette conclusion n'est pas plausible, car les incohérences entre les sources vont bien au-delà des différences de moyennes. Au total, 195 des 435 clients inclus dans les résultats ont affirmé n'avoir jamais emprunté à l'IMF. Or, cette explication de la « honte du crédit » pour ces ménages induirait une explication de « fierté du crédit » pour les 152 ménages ayant déclaré avoir obtenu un prêt de l'IMF sans figurer dans ses registres.

Par ailleurs, selon les données d'enquête recueillies sur l'échantillon du panel, l'accès au crédit est resté stable dans le groupe de traitement entre l'enquête initiale et l'enquête finale, alors qu'il a diminué dans le groupe de contrôle (la microfinance marocaine a traversé une crise majeure de 2008 à 2010). Nos résultats remettent en question le sens même de cette RCT : ce qui a été testé ne semble pas être l'impact de l'introduction du microcrédit dans des zones « vierges », mais plutôt le remplacement d'autres sources formelles par une source de microcrédit dans le groupe de traitement et un rationnement du crédit dans le groupe de contrôle.

Nous avons également trouvé des erreurs d'échantillonnage. Par exemple, la composition par sexe et par âge pour 20 % des ménages interrogés dans l'enquête initiale et réinterrogés dans l'enquête finale diffère à tel point qu'il est peu probable que ce soient les mêmes unités qui aient été réinterrogées. En outre, nous avons constaté que les caractéristiques de l'échantillon de Crépon *et al.* différaient sensiblement de celles de la population. La composition des ménages est passée de 5,17 à 6,13 membres entre l'enquête initiale et l'enquête finale. Or, le recensement national indique que les ménages ruraux marocains comptaient en moyenne 6,03 membres en 2004 et 5,35 membres en 2014. De telles divergences soulèvent des questions sur la représentativité de l'échantillon et compromettent donc la validité externe de cette étude.

Les auteurs ont produit une réponse à notre réplique, intitulée *Rejoinder*, réfutant la plupart des erreurs que nous avons documentées (CRÉPON *et al.*, 2019). Ils se sont référés à notre analyse, mais ne semblent pas avoir répliqué ou analysé de près son contenu statistique. Par ailleurs, le *Rejoinder* contient

de nombreuses erreurs et omissions factuelles. Nous avons publié un document passant en revue leurs principaux arguments en réponse à notre réplique (BÉDÉCARRATS *et al.*, 2019c). Nous avons constaté que toutes les erreurs de codage, de mesure et d'échantillonnage documentées dans notre réplique demeurent d'actualité.

Distorsion du protocole : ajustement du produit et de l'échantillonnage

Notre deuxième article a cherché à expliquer comment de telles incohérences pouvaient se produire, à l'appui d'une étude qualitative de terrain spécialement conçue pour compléter la RCT (MORVANT-ROUX *et al.*, 2014) et de différents documents et données publics et internes des principaux acteurs de la RCT (BÉDÉCARRATS *et al.*, 2019b). L'article décrit l'ensemble de la chaîne de production de l'étude, depuis l'échantillonnage, la collecte, la saisie et le recodage des données à la publication et à la diffusion des résultats en passant par les estimations et interprétations. Loin des conditions idéales de laboratoire⁹, l'analyse de la mise en œuvre du protocole randomisé sur le terrain par les différents acteurs (chacun ayant ses propres motivations et contraintes) révèle un certain nombre de disparités par rapport au protocole théorique annoncé dans l'article publié.

L'un des principaux problèmes rencontrés durant l'étude a été le taux d'adhésion, bien inférieur aux prévisions initiales, ce qui a nécessité un certain nombre de mesures correctives. Le premier ajustement a consisté à modifier l'intervention (offre de microcrédit) en lançant de nouvelles campagnes d'information, en accordant des primes forfaitaires aux agents et en éliminant le quota minimum de femmes. L'adhésion est devenue une « obsession », tant pour l'équipe de recherche que pour les agents de crédit, qui ont eux-mêmes employé ce terme et se sont donné beaucoup de mal pour inciter les villageois à souscrire un microcrédit. Parmi les stratégies utilisées, il a été décidé de repousser les frontières habituelles des villages dans l'espoir de trouver davantage de clients¹⁰. Lorsque ces mesures se sont révélées insuffisantes, l'équipe a ajusté la méthode d'échantillonnage (modification des modèles prédictifs et ajout dans l'enquête finale de nouveaux ménages, censés présenter une plus grande propension à emprunter). Les villages affichant un taux d'adhésion nul ont été éliminés.

Mauvaise qualité des données et erreurs de mesure

La collecte et la saisie des données ont été sous-traitées à une société de conseil spécialisée dans l'ingénierie, mais sans aucune expérience des enquêtes statistiques. Afin de suivre la conception et la mise en œuvre de la RCT, le bailleur de fonds de la RCT (l'Agence française de développement – AFD) a nommé une

9. Les expérimentations de terrain telles que les RCT sont précisément conçues pour sortir du monde artificiel des laboratoires. Or, les *randomistas* ont trop souvent tendance à penser que le protocole peut être appliqué tel quel, comme en laboratoire, ce qui n'est pas le cas.

10. Un changement de produit pour l'intervention évaluée par la RCT pose également un problème de validité externe (car les conditions de l'expérimentation ne sont pas conformes à son fonctionnement dans le « monde réel ») (PETERS *et al.*, 2018).

équipe d'économistes et de spécialistes des enquêtes auprès des ménages. Lors de missions de terrain, cette équipe a très tôt constaté de graves dysfonctionnements dans la collecte des données. Ils ont observé des problèmes de traduction, car les enquêteurs ne parlaient pas le berbère, langue parlée par une grande partie de la population cible. Les enquêteurs ont donc eu largement recours à des traducteurs improvisés, parmi lesquels des dirigeants locaux, ce qui a induit des problèmes de compréhension et de biais dans les réponses (désirabilité sociale et méfiance à l'égard du gouvernement).

Un autre dysfonctionnement portait sur le nombre de répondants dans les ménages et les familles élargies, qui semblait ici encore relever de la pure improvisation, en fonction de la présence et de la disponibilité des personnes, mais aussi de leur capacité à se comprendre entre elles et à comprendre les enquêteurs. Ces observations expliquent probablement en partie les importantes divergences susmentionnées entre l'enquête initiale et l'enquête finale. Toutefois, l'ampleur des divergences suggère une autre explication : certains ménages n'étaient peut-être tout simplement pas les mêmes, comme le confirme notre réplique. L'absence d'adresse précise exige des techniques de suivi rigoureuses, qui ont peut-être été négligées. Faute de temps et de supervision, il est possible que certains enquêteurs se soient contentés d'interroger les ménages disponibles au moment de leur visite. L'équipe de l'AFD a formulé des recommandations pour améliorer la qualité des données collectées, en s'inquiétant des répercussions potentielles de ces lacunes sur les résultats de l'expérience. Elle a également dénoncé les problèmes de saisie des données. Bien que l'équipe du J-PAL ait répondu, contestant la gravité des problèmes et affirmant qu'ils ne remettaient pas en cause la validité interne de l'expérience, le comité de pilotage a décidé, lors de sa réunion suivante, que tous les questionnaires déjà saisis devaient être envoyés à l'Institut national français de la statistique (Insee) à Paris pour être à nouveau saisis.

Ces différents points n'ont pas été évoqués dans l'article publié et révèlent des lacunes dans la préparation, la mise en œuvre et le suivi du travail sur le terrain.

Au-delà de la RCT marocaine : évaluation générale

Il n'est pas possible d'analyser les cinq autres RCT du numéro spécial de manière aussi détaillée, aussi bien pour des raisons de temps que parce qu'il manque les données brutes pour deux d'entre elles (tabl. 3). Nous nous livrons donc à un exercice partiel, à savoir une lecture critique des principales caractéristiques techniques des procédures employées, en nous basant sur les informations disponibles dans les articles publiés. Le tabl. 4 résume les problèmes de validité interne tels qu'ils peuvent être évalués à partir des informations dont nous disposons. Presque aucun de ces problèmes n'est évoqué dans le numéro spécial, et encore moins dans l'introduction générale. Nous abordons tour à tour la question de l'erreur d'échantillonnage et la question de l'inexactitude de mesure.

Tableau 3
Validité externe, réserves reconnues et préoccupations éthiques.

	Bosnie- Herzégovine	Éthiopie	Inde	Mexique	Mongolie	Maroc
Population d'intérêt	Expansion de l'IMF	Expansion de l'IMF	Expansion de l'IFM (partielle)	Expansion de l'IMF	Expansion de l'IMF	Expansion de l'IFM (partielle)
Extrapolation à population plus générale ?	Non	Non	Non	Non	Non	Non
Discussion sur les limites potentielles (dans l'article) ?						
Hawthorne ou John Henry	Oui	Non	Non	Non	Non	Non
Équilibre général	Non	Non	Non	Non	Non	Oui
Comparaison avec des données d'OSN ?	Oui	Non	Non	Non	Oui	Non
Autres enquêtes/méthodes mises en œuvre ?	Non	Non	Non	Non	Enquêtes dans les villages, entretiens qualitatifs	Non
Si oui, utilisées ?	–	–	–	–	Non	–
Réserves explicites reconnues ?	Oui	Oui	Oui	Oui	Oui	Oui
	1-Pas de validité externe 2-Sous-puissance statistique 3-Effets potentiels H & JH	1-Pas de validité externe 2-Sous-puissance statistique 3-Pas de panel = déséquilibre au niveau de l'EI, attrition sélective, effet hétérogène 4-Non-respect du plan expérimental 5-Pas de consommation 6-Erreurs de mesure	1-Sous-puissance statistique 2-EI non représentative 3-Attrition sélective et migration 4-Contamination 5-IdT représentative des « emprunteurs probables » uniquement	1-Pas de validité externe 2-Qualité des données 3-Pas d'EI 4-Périodes de traitement hétérogènes	1-Pas de validité externe 2-Sous-puissance statistique 3-Présence d'autres IMF 4-Attrition (déséquilibre possible) 5-Non robuste au TMH	1-Petits déséquilibres significatifs au niveau de l'EI

	Bosnie- Herzégovine	Éthiopie	Inde	Mexique	Mongolie	Maroc
Discussion sur les préoccupations éthiques						
Consentement éclairé à l'expérimentation	Non	Non	Non	Non	Non	Non
Analyse et suivi des risques	Non	Non	Non	Non	Non	Non
Équipoise	Non	Non		Non	Non	Non
Reproductibilité						
Données disponibles	Données brutes	Non	Données agrégées	Données agrégées	Données brutes	Données brutes
Code détaillé disponible	Oui	Non	Partiellement	Partiellement	Oui	Oui
Questionnaire d'enquête disponible sur le site d'AEJ	Oui	Non	Non	Non	Non	Oui

Source : Florent Bédécarrats, Isabelle Guérin et François Roubaud, sur la base de BANERJEE *et al.* (2015c).

Notes : IMF : institution de microfinance ; OSN : Office statistique national ; IdT : intention de traiter ; EI : enquête initiale ; EF : enquête finale ; H&JH : Hawthorne et John Henry ; TMH : test multi-hypothèses.

Tableau 4
Validité interne des six RCT.

	Bosnie-Herzégovine	Éthiopie	Inde	Mexique	Mongolie	Maroc
Population d'intérêt	Clients potentiels initialement refusés par l'IMF pour cause d'insolvabilité	Ménages ruraux dans deux zones ad hoc	Emprunteurs potentiels (femmes vivant dans des bidonvilles depuis plus de trois ans avec une pièce d'identité valide) dans les zones d'expansion de l'IMF à Hyderabad	Clients potentiels (femmes possédant ou envisageant de créer une entreprise ou ayant l'intention d'emprunter) dans les zones d'expansion de l'IMF à Central Sonora, au Mexique	Femmes pauvres : (actifs < 869 \$ et bénéfiques < 174 \$/mois) Inscrites pour obtenir un prêt	Ménages à forte propension à l'emprunt dans les zones rurales d'extension de l'IMF
Plan de sondage, randomisation						
Plan de sondage	Échantillon individuel à choix raisonné	– Stratifié (2 « zones ») – 3 degrés (unités administratives/village/ménage)	2 degrés (bidonvilles/ménages)	2 degrés (village/ménage)	– Stratifié (5 provinces) – 2 degrés (village/ménage) Mongolie du Nord	2 degrés (village/ménage)
Informations sur la sélection des zones	Non applicable	Oui	Oui	Oui	Non	Oui
Nombre de zones (T, C)		353 villages	104 (52, 52)	238 (120, 118)	25 (15, 10)	162 (81,81)
Zones abandonnées	Non applicable	Non	Oui (16 bidonvilles)	Oui (12 zones)	Non	Oui (non précisé)
Informations sur la sélection des personnes	Oui, non aléatoire	Oui, aléatoire	Oui, aléatoire	Oui, aléatoire	Oui, non aléatoire (30 premiers à s'inscrire)	Oui, aléatoire
Informations sur la randomisation (T vs C)	Oui (niveau individuel)	Oui (niveau du village)	Oui (niveau du bidonville)	Oui (niveau de la zone)	Oui (niveau du village)	Oui (niveau du village)

	Bosnie-Herzégovine	Éthiopie	Inde	Mexique	Mongolie	Maroc
Taille de l'échantillon (complet ; contrôle)	EI (1 196 ; 568) EF (994 ; 444)	EI (6 412 ; n.d.) EF (6 263 ; n.d.)	EI (2 800 ; 1 220) EF1 (6 863 ; 3 264) EF2 (6 142 ; 2 943)	EI (6 786 ; n.d.) EF (16 560 ; 8 298)	EI (710 ; 299) EF (610 ; 260)	EI (4 465 ; 2 266) EF (5 551 ; 2 810)
Taux d'attrition (EI->EF) : % total, % contrôle	Panel (17 % ; 22 %)	Pas de panel 2 groupes types	EI->EF : pas de panel EF1->EF2 (11 % ; 10 %)	Panel (37 % ; n.d.)	Panel (16 % ; 15 %)	Panel (8 % ; 7 %)
Respect du protocole expérimental	Oui	Non 22 % de zones mal allouées (12 % T non traités, 23 % C traités)	Non (16 zones supprimées ; EF non fiable)	Non (EI avortée)	Oui	Non (nouveaux ménages ajoutés à EF)
Tests d'équilibre dans l'enquête initiale						
Population incluse	Ménages du panel uniquement	Ménages du panel uniquement	Tous les ménages EI	Ménages du panel uniquement	Ménages du panel uniquement	Tous les ménages EI
Variables testées	27	35	33	14	48	43
Inclusion des principaux résultats de l'étude	Oui	Oui	Oui	Non	Oui	Non
Déséquilibres importants signalés	Non	Non	Non	Oui	Oui	Oui
Trimming	Résultats avec et sans <i>trimming</i> de 1 % pour les contrôles de robustesse	Résultats avec et sans <i>trimming</i> 8 obs. pour les contrôles de robustesse	Non	Non	Non	EI : <i>trimming</i> des valeurs les plus élevées pour 10,3 % des obs. EF : <i>trimming</i> de 0,5 % des obs.
Qualité des données (discussion dans l'article)	Non	Oui, marginale (erreurs de mesure)	Oui, marginale (éventuelles erreurs de mémoire)	Oui, marginale (variables de résultats manquantes pour EF)	Non	Non (excepté pour l'adhésion) (données administratives vs enquête)

Source : Florent Bédécarrats, Isabelle Guérin et François Roubaud, sur la base de ANGELUCCI *et al.*, 2015 ; ATTANASIO *et al.*, 2015 ; AUGSBURG *et al.*, 2015 ; BANERJEE *et al.*, 2015b ; CRÉPON *et al.*, 2015 ; TAROZZI *et al.*, 2015.

Note : EI : enquête initiale ; EF : enquête finale ; n.d. : non disponible ; obs. : observations ; T vs C : groupe de traitement par rapport au groupe de contrôle.

Concernant l'échantillonnage, il convient de noter que, de manière générale, les articles ne fournissent pas les éléments de base permettant de décrire et de qualifier avec précision les plans d'échantillonnage et de sélection adoptés (il existe pourtant des normes standardisées pour de telles descriptions, comme dans Statistique Canada, 2010 et ARDILLY et TILLÉ, 2006). Les auteurs concentrent leurs analyses sur les questions de randomisation et d'inférence causale. Premièrement, la population de référence n'est jamais clairement établie. Dans la plupart des cas, elle correspond aux clients éligibles dans les zones d'expansion de l'IMF, quoique l'on ne sache pas comment ces dernières sont définies. Cela a des répercussions problématiques sur la validité externe des RCT (voir la partie « Validité externe » ci-après). Deuxièmement, les plans d'échantillonnage adoptés s'inscrivent dans la catégorie générale de l'échantillonnage aléatoire stratifié à plusieurs degrés, à l'exception des RCT en Bosnie-Herzégovine et en Mongolie. Aucun de ces deux cas ne fait en effet l'objet d'un échantillonnage aléatoire : en Mongolie, les trente premières femmes pauvres dans chaque village sélectionné à se dire intéressées par un prêt ont été retenues ; en Bosnie-Herzégovine, il a été demandé aux agents de crédit de sélectionner des clients potentiels qui n'étaient pas jugés éligibles selon les normes actuelles de l'IMF. Dans tous les cas, ces plans de sondage complexes, pour reprendre la terminologie statistique, soit ne permettent pas de calculer les intervalles de confiance associés à l'impact estimé (les deux cas mentionnés ci-dessus), soit nécessiteraient des calculs particulièrement complexes pour estimer la variance, calculs qui ne sont pas faits (sauf pour l'estimation des erreurs types robustes propres aux groupes). La conséquence directe de cette divergence est que les intervalles de confiance sont probablement sous-estimés et que les impacts réputés significatifs, déjà peu nombreux, ne devraient pas être statistiquement différents de zéro.

De plus, quatre des six RCT se sont écartées du protocole canonique de la méthode expérimentale : sélection aléatoire d'un groupe de traitement et d'un groupe de contrôle, enquête initiale avant traitement, puis suivi du panel basé sur une enquête finale après traitement. Dans le cas de l'Éthiopie, les enquêtes initiale et finale n'ont pas porté sur des panels, mais sur des enquêtes en coupe répétées (c'est-à-dire que différents individus ont été interrogés). Il apparaît donc impossible d'identifier dans l'enquête initiale les déséquilibres potentiels concernant la population pour laquelle l'impact est estimé dans l'enquête finale. Dans les cas du Mexique, du Maroc et de l'Inde, les enquêtes sur le terrain n'ont pas pu être menées comme prévu initialement et les risques pesant sur la réussite de l'expérience ont conduit à réajuster le protocole initial en cours de route. En Inde, l'enquête initiale n'a servi de panel de base pour aucune des deux enquêtes finales subséquentes¹¹, soulevant les mêmes problèmes que dans le cas de l'Éthiopie mentionné ci-dessus. Au Mexique, l'enquête initiale a avorté en raison de la mauvaise qualité des données recueillies : 73 % des

11. BANERJEE *et al.* (2015b) ont conduit une première enquête finale en 2007 et 2008. Ils ont réinterrogé les ménages de la première enquête finale dans le cadre d'une deuxième enquête finale en 2009 et 2010.

ménages de l'enquête initiale n'ont pas été revus pour l'enquête finale et 89 % de l'échantillon final n'avaient pas été interrogés lors de l'enquête initiale, de sorte que la majorité des ménages interrogés pour l'enquête finale ont été ajoutés à ce stade. Une stratégie similaire a été adoptée au Maroc. En raison de la faible adhésion parmi les ménages identifiés comme emprunteurs potentiels, de nouveaux ménages ont été sélectionnés dans l'enquête finale, représentant 26 % de l'échantillon final. Si l'on tient également compte des taux d'attrition (disponibles uniquement pour les protocoles des panels) compris entre 8 % (Maroc) et 37 % (Mexique), il est clair qu'aucune des RCT n'a été menée conformément aux normes (échantillonnage non aléatoire des ménages ciblés en Bosnie-Herzégovine et en Mongolie, absence ou échec des panels pour les quatre autres RCT en raison de problèmes de collecte de données ou d'une faible adhésion).

Quoi qu'il en soit, il aurait été fondamental de vérifier l'équilibre des échantillons sur l'enquête initiale. Les études varient considérablement en termes de variables testées. Certaines ont testé un nombre étonnamment réduit de variables par rapport au large éventail de données collectées (Mexique). D'autres en ont testé bien davantage, mais toutes diffèrent quant aux variables testées. Dans certains cas, la plupart des variables incluent au moins quelques-uns des résultats pour lesquels l'impact a été mesuré dans l'enquête finale. Dans le cas du Maroc, toutefois, les tests d'équilibre n'ont été appliqués qu'à des sous-ensembles spécifiques des variables de résultats évaluées dans l'enquête finale (par exemple, les ventes pour les ménages de cultivateurs ou d'éleveurs de bétail, en lieu et place des ventes globales déclarées dans l'enquête finale). Dans notre réplique, nous avons constaté des déséquilibres importants et significatifs dans ces résultats. Les ménages du groupe de traitement ont tiré 22 % moins de ventes et de bénéfices de leur travail indépendant que les ménages du groupe de contrôle (significatif au niveau de 5 %). Ils ont également investi 61 % de plus (significatif au niveau de 5 %). En outre, il existe dans l'enquête initiale des déséquilibres concernant un certain nombre de variables importantes, telles que la superficie des terres possédées, l'accès aux services de base et l'autonomisation des femmes. Outre les variables testées, la base de calcul a aussi son importance. Par exemple, l'étude mexicaine a limité ses tests d'équilibre à 1 823 ménages interrogés, tant dans l'enquête initiale que dans l'enquête finale. Or, si l'on pratique les mêmes tests sur tous les ménages enquêtés en *baseline* (6 786), comme cela a été fait en Inde et au Maroc, on constate des différences significatives dans le revenu des ménages par adulte au cours du mois précédent, en particulier chez ceux faisant partie d'un groupe informel¹².

Même si les différences de base entre les groupes de traitement et de contrôle ne sont pas statistiquement significatives, elles peuvent être considérables. En Mongolie et en Éthiopie, les tests d'équilibre en enquête initiale ont révélé des différences moyennes souvent supérieures à 10 % (jusqu'à 50 %), mais non significatives (ce qui n'est pas surprenant étant donné la petite taille des

12. Les calculs sont disponibles sur demande auprès des auteurs.

échantillons). Ils sont systématiquement interprétés au nom de ce qui semble être un souci de commodité (absence de déséquilibres et donc succès du processus de randomisation), alors que l'explication inverse est souvent avancée pour les résultats : lorsque les coefficients ne sont pas significatifs en raison du manque de puissance statistique, ils sont interprétés comme étant « économiquement significatifs ».

Aucun des articles n'aborde en profondeur les questions d'erreur de mesure. Pourtant, la littérature insiste sur la difficulté d'obtenir des mesures fiables pour bon nombre des résultats analysés, en particulier la consommation des ménages et la production des micro-entreprises et de l'agriculture (DEATON, 1997 ; GROSH et GLEWWE, 2000). Les erreurs de mesure sont simplement mentionnées dans une note de bas de page consacrée au biais de mémoire potentiel dans le cas de l'Inde et dans une discussion sur la sous-déclaration des emprunts dans le cas du Maroc, supposée expliquer les différences entre les données administratives et les enquêtes. Seule la RCT éthiopienne émet de sérieux doutes sur la qualité des données et reconnaît explicitement que ce problème affecte la validité interne. La RCT mexicaine précise que l'enquête initiale a dû être interrompue et que ses données n'ont pas pu être utilisées parce qu'elles n'étaient pas fiables, sans donner de détails ni expliquer comment des données plus fiables auraient pu être recueillies dans l'enquête finale. Malheureusement, il apparaît impossible de traiter plus en détail la qualité des données à partir des seuls articles. Cependant, une analyse détaillée de la cohérence des données et du recodage pratiqué par des chercheurs dans le cas marocain (BÉDÉCARRATS *et al.*, 2019a) montre que ce problème a altéré les résultats. Certains éléments tendent à indiquer que des problèmes similaires peuvent tout à fait se présenter dans d'autres cas. Par exemple, une analyse préliminaire des données mexicaines révèle que les tranches d'âge ne correspondent pas entre les enquêtes pour 231 (12,7 %) des 1 823 femmes interrogées, à la fois dans l'enquête initiale et dans l'enquête finale de BANERJEE *et al.* (2015b).

Validité externe

La question de la validité externe des RCT est la plus débattue dans la littérature sur ce sujet. La validité externe est une question primordiale, d'autant plus que, contrairement à beaucoup de données observationnelles, les RCT sont menées à petite échelle et dans des lieux non représentatifs, comme on l'a vu plus haut. La validité externe est également menacée lorsque l'échantillonnage est sélectif, c'est-à-dire lorsqu'une étude se concentre sur des sites et des catégories de population spécifiques. Intervient ensuite le biais lié aux opérateurs, par exemple, lorsque les résultats obtenus par une ONG diffèrent d'une même intervention réalisée à plus grande échelle par un gouvernement (BOLD *et al.*, 2013 ; VIVALD, 2020). Or, la question de la validité externe des RCT est rarement prise en compte par les *randomistas*. PETERS *et al.* (2018) ont systématiquement passé en revue toutes les (54) RCT publiées dans les principales revues économiques de 2009 à 2014 afin d'évaluer les menaces les plus préoccupantes pesant sur la validité externe (effets

Hawthorn/Henry¹³, effets d'équilibre général, problèmes d'échantillons spécifiques et attention accordée à l'offre de traitement). Sur la base d'un ensemble d'indicateurs objectifs, mais de critères statistiquement moins exigeants, l'article constate que la majorité des RCT publiées passent ces menaces sous silence et que beaucoup ne donnent pas les informations nécessaires pour évaluer les problèmes potentiels.

La validité externe a aussi un lien avec la pertinence des résultats sélectionnés. Se concentrer sur l'impact « moyen » et la difficulté à traduire l'hétérogénéité des impacts et leur distribution constituent un obstacle majeur à la pertinence des résultats (RAVALLION, 2009a ; DFID, 2012 ; VIVALD, 2020). Le fait de se cantonner à un impact à court terme (pour des raisons de coût et d'attrition) signifie souvent que l'on étudie des indicateurs à mi-parcours, qui peuvent se révéler très différents des résultats finaux (BOONE *et al.*, 2013), si ce n'est totalement inverses, puisque les trajectoires de projets sont rarement linéaires (LABROUSSE, 2010 ; WOOLCOCK, 2013). Pourtant nombreux, les effets d'entraînement et d'équilibre général sont occultés, même si ce n'est que partiellement le cas pour la RCT marocaine (RAVALLION, 2009a ; ACEMOGLU, 2010 ; DEATON et CARTWRIGHT, 2018). Il en va de même pour les considérations politiques afférentes à la réplification des programmes, en dépit de leur importance pour la transposition à plus grande échelle (ACEMOGLU, 2010 ; BOLD *et al.*, 2013 ; PRITCHETT et SANDEFUR, 2013b). Enfin, point essentiel, les *raisons* de l'impact sont ignorées : les RCT peuvent fort bien permettre de mesurer et de tester certains impacts et aspects des interventions, mais elles ne peuvent analyser ni leurs *mécanismes* ni leurs *processus* sous-jacents. Nonobstant les limites de la méthode, l'absence de théorie empêche toute forme de compréhension des processus de changement. Pour pallier cette limitation de la théorie probabiliste de la causalité, il faudrait un « modèle causal » (CARTWRIGHT, 2010), une théorie cohérente du changement (WOOLCOCK, 2013), une approche structurale (ACEMOGLU, 2010) et une évaluation de l'intervention dans son contexte (RAVALLION, 2009a ; PRITCHETT et SANDEFUR, 2015).

Le tabl. 3 résume les problèmes de validité interne tels qu'ils peuvent être évalués à l'appui des informations dont nous disposons. Les lacunes usuelles des RCT demeurent valables ici.

En premier lieu, l'échantillonnage est sélectif : les critères de sélection de l'expérience sont *ad hoc*, puisque les RCT ont été conduites dans des zones d'expansion des IMF. Comme le démontre WYDICK (2016), la contrainte de la randomisation (identification de populations ou de zones vierges) a obligé les *randomistas* à choisir des populations et des zones « marginales » jusque-là négligées par les IMF et donc très spécifiques par rapport au marché « normal ». Les tentatives infructueuses menées par les études marocaines, mexicaines et indiennes pour

13. Il s'agit de biais comportementaux induits par l'expérience dès lors que les sujets savent qu'ils y participent : biais sur le groupe de traitement (effet Hawthorne) ou sur le groupe de contrôle (effet John Henry). Dans le domaine médical, les RCT en simple ou double aveugle (sujets et expérimentateurs) sont ordinairement utilisées pour contrôler ces biais (voir Abramowicz et Szafarz, chap. 10, ce volume).

identifier des emprunteurs potentiels montrent qu'il est difficile de caractériser la population cible du microcrédit. Cela exclut en toute légitimité la possibilité d'extrapolation à une population plus large. *A fortiori*, les échantillons retenus ne sont représentatifs de rien, si ce n'est d'eux-mêmes : les ménages interrogés dans le cas de la Bosnie-Herzégovine et de la Mongolie, et les zones d'expansion (villages et quartiers sélectionnés) dans le cas des quatre autres pays. De plus, cette propriété n'existe qu'en théorie : les multiples défaillances des protocoles d'enquête sur le terrain font que les échantillons théoriquement représentatifs des zones d'expansion ne le sont pas *de facto*.

Si les données ne peuvent pas être extrapolées, la comparaison avec d'autres sources peut se révéler utile pour qualifier les profils des personnes interrogées. Les chiffres officiels provenant d'enquêtes représentatives conduites par des offices nationaux des statistiques constituent un bon point de repère pour caractériser un contexte national ou local. Seules deux études l'ont fait (Bosnie-Herzégovine et Mongolie). Dans les quatre autres études, il est difficile de savoir qui sont les personnes interrogées. Comme indiqué ci-dessus, nous avons effectué cet exercice pour la RCT marocaine. Nous avons démontré, entre autres résultats, que la taille moyenne des ménages est atypique et tend à augmenter, alors qu'elle décroît dans le reste de la population sur la même période. Pour aller plus loin dans cette évaluation, nous utilisons la typologie des risques sur la validité externe établie par PETERS *et al.*, (2018) : effets Hawthorne et John Henry et effets d'équilibre général (les autres étant abordés ci-dessus). Les articles ne traitent pas de ces risques et beaucoup ne fournissent pas les informations nécessaires pour évaluer les problèmes potentiels, à l'exception (partielle) des effets Hawthorne (Bosnie-Herzégovine et indirectement Mexique, voir la discussion sur l'éthique ci-dessous) et des effets d'équilibre général et d'entraînement (Maroc), bien que ces effets soient à l'œuvre dans tous les cas.

Ces risques pour la validité externe et interne sont-ils reconnus par les *randomistas* ? Plus généralement, quels types de réserves formulent-ils dans leurs articles ? Nous en rendons compte dans le tabl. 3. À l'exception de la RCT marocaine, les auteurs formulent un certain nombre de réserves. Presque tous mentionnent le manque de validité externe induit par le manque de puissance statistique, lui-même dû à la taille insuffisante des échantillons. De même, l'hétérogénéité des traitements est largement reconnue. Le fait que les autres RCT donnent des résultats similaires (mais tout aussi insuffisants en termes de puissance statistique) est perçu comme une source de robustesse (voir, par exemple, BANERJEE *et al.*, 2015b : 25). En outre, des réserves plus spécifiques sont citées, notamment au regard du non-respect du plan de sondage (Éthiopie), de l'attrition sélective (Inde et Éthiopie) et des erreurs de mesure (Éthiopie). Ces observations tendent à confirmer la persistance des résultats de PETERS (2018) concernant l'attention limitée accordée à la validité externe, ce à quoi il faut ajouter les problèmes de validité interne évoqués ci-dessus.

Enfin, les considérations éthiques méritent d'être discutées, car elles revêtent une dimension spécifique pour les RCT en général (voir l'introduction, ce volume ; Ravallion, chap. 1, ce volume ; et Abramowicz et Szafarz, chap. 10,

ce volume). Or, contrairement à toute attente, aucun des articles ne fait état de ces considérations. Ils ne précisent pas si le consentement éclairé des participants a été demandé et obtenu, à l'exception d'ANGELUCCI *et al.* (2015) sur Compartamos au Mexique. En outre, les informations qu'ils déclarent avoir communiquées aux participants sont partielles : ils précisent, peut-être pour écarter tout soupçon d'effet Hawthorne, qu'ils leur ont demandé leur accord pour participer à une « enquête de recherche socio-économique approfondie ». Mais ils ont sciemment omis de mentionner que l'enquête était en lien avec Compartamos et, surtout, qu'elle faisait partie d'une expérimentation. L'examen des questionnaires d'enquête disponibles (Bosnie-Herzégovine et Maroc) montre que, dans ces deux cas, les répondants n'ont pas été informés qu'ils participaient à une expérimentation. La RCT bosniaque soulève d'autres questions éthiques. Cette RCT consistait à accorder des crédits à des personnes qui, selon les critères de solvabilité de l'IMF, avaient initialement été refusées, comme en Afrique du Sud et aux Philippines (KARLAN et ZINMAN, 2009 ; 2011). Cette stratégie faisait peser un risque sur le groupe traité, au mépris du principe de « ne pas nuire ». La RCT confirme que les clients marginaux ont beaucoup plus de mal à rembourser que les clients réguliers, avec un risque de surendettement¹⁴.

Maintenant que nous avons discuté des questions de validité interne et externe, penchons-nous sur la question des impacts eux-mêmes. Même si l'on ne prend pas en compte les limites exposées ci-dessus et que l'on s'en tient aux résultats proposés par les auteurs, les impacts sont problématiques. Le tabl. 5 en donne un aperçu. Premièrement, les données sur l'adhésion ne sont pas fiables et présentent souvent des contradictions entre les sources d'enquête et les sources administratives. Le cas marocain montre que les incohérences vont au-delà des différences de moyennes et de la sous-déclaration (voir la section « Le cas emblématique de la RCT marocaine » sur les écarts entre les données administratives et les données d'enquête). En moyenne, l'impact des expériences sur l'adhésion du crédit varie de 8 à 50 %, lorsque les groupes ont été randomisés, à 98,5 % en Bosnie-Herzégovine, où ce sont les individus qui ont été randomisés.

En ce qui concerne les impacts sur le microcrédit, le faible taux d'adhésion a de considérables implications sur le niveau de significativité des coefficients estimés. DAHAL et FIALA (2020) répliquent les six RCT de l'*AEJ:AE*. Ils constatent que chacune d'entre elles souffre d'un important manque de puissance statistique en raison de la faible adhésion du produit financier offert. Même après la mise en commun des données, les magnitudes minimales des effets détectables demeurent très élevées : 230 % pour les principaux résultats en conformité parfaite et 1 000 % en conformité effective. Ils concluent dans leur résumé que « les études existantes sur l'impact de la microfinance pâtissent généralement d'un manque de puissance statistique pour identifier les impacts de manière fiable et suggèrent que nous en savons encore très peu sur l'impact de la microfinance ». Quoique BANERJEE *et al.* (2015b) admettent le problème de la sous-puissance statistique

14. « Tout cela indique que les agents de crédit avaient de bonnes raisons de classer comme marginale notre population cible » (AUGSBURG *et al.*, 2015 : 201).

dans leur introduction, l'article de DAHAL et FIALA (2020) est le premier à quantifier l'ampleur du problème. Il confirme l'étude précédente de MCKENZIE (2012), qui estime à 15 000 000 la taille d'échantillon nécessaire pour pouvoir garantir la capacité à déceler des magnitudes d'impact de 10 % dans la RCT indienne.

Pour ce qui est des impacts sur les résultats sélectionnés, la présentation faite à ce sujet par les auteurs de l'introduction générale (tabl. 2), qui est supposée résumer les résultats consolidés des six RCT, est trompeuse. Un décompte exhaustif des impacts estimés sur toutes les variables examinées dans les six articles permet de tirer les conclusions suivantes. Pas moins de 298 impacts sont estimés sur l'ensemble du volume (sans compter les estimations des quantiles). Sur ce total, seuls 10 sont significatifs au niveau de 1 %, ce qui signifie que 97 % des effets possibles retenus ne sont pas significativement différents de zéro. Trois RCT n'ont aucun impact significatif (Bosnie-Herzégovine : 0/47, Éthiopie : 0/37 et Mongolie : 0/41) et l'une d'entre elles n'a qu'un seul impact significatif (Inde : 1/99). Même lorsque le seuil est abaissé à 10 % (un niveau moins rigoureux que dans la pratique habituelle), 81 % des effets ne sont pas significatifs. La RCT bosnienne est un cas extrême à cet égard, avec seulement trois impacts significatifs à ce seuil sur les 47 testés. Ces proportions suscitent d'autant plus de doutes que tous les articles dénoncent un problème systématique de sous-puissance statistique, ce qui expliquerait l'absence d'impact. La taille des échantillons n'est pas suffisante pour estimer les impacts, compte tenu du faible taux d'adhésion, et c'est effectivement ce que nous constatons. De plus, 60 % des impacts significatifs (à 1 %) proviennent de la RCT marocaine, alors qu'elle représente à peine 12 % du nombre total d'impacts estimés. Ce résultat confirme le rôle central joué par cette expérience dans le numéro spécial, indépendamment de ses vertus supposées concernant sa stratégie novatrice d'échantillonnage et sa tentative pionnière d'estimation des effets d'entraînement. Cependant, nous avons montré le caractère douteux des résultats obtenus par cette RCT. Cela réduit encore le nombre d'impacts significatifs, qui était déjà notablement faible.

Symptomatiquement, le cheminement des résultats des articles académiques, d'abord dans l'introduction générale, puis dans la synthèse proposée dans le *Policy Bulletin* (J-PAL et IPA, 2015) participe, par des approximations successives, de la simplification et de la généralisation abusive des leçons retenues, voire de la présentation de résultats erronés. Si nous revenons au résumé des impacts présenté dans le *Policy Bulletin* (p. 11, tabl. 2 ; voir aussi notre tabl. 2), sur les 48 impacts mesurés (8 résultats et 6 pays), 16 sont annoncés comme significatifs (14 positifs et 2 négatifs). On est loin du compte. Premièrement, le seuil de significativité choisi est de 10 %, ce qui correspond à un niveau de précision situé à la limite supérieure de celui qui est habituellement utilisé. Si nous adoptons un seuil plus exigeant et plus proche des pratiques habituelles (à savoir 1 %), aucun des 16 impacts n'est significatif.

Une analyse plus détaillée des 16 impacts sélectionnés révèle de nombreuses incohérences. Pour la Bosnie-Herzégovine, les impacts sur la *propriété des entreprises* et sur les *stocks/actifs des entreprises* sont annoncés comme positifs. Mais le premier n'est pas significatif à 10 %. Quant au second, ce qui est

significatif à 10 % est une variable fictive qui mesure si l'entreprise détient ou non du capital. L'impact sur la valeur totale des *actifs* est négatif (quoique non significatif), donc au mieux nul. Pour l'Éthiopie, le seul impact considéré comme significatif et négatif est celui sur les *dépenses/consommation des ménages*. Or, la consommation n'a pas été mesurée dans l'enquête. En Inde, les deux impacts positifs sont sur les *stocks/actifs des entreprises* et les *stocks/coûts des entreprises*. Aucun de ces deux impacts n'est robuste : le premier impact est positif dans la deuxième enquête finale, mais n'est pas significatif dans l'enquête initiale, et inversement pour l'impact sur les *stocks/coûts des entreprises*. Au Mexique, on note deux impacts positifs. Si l'impact se vérifie pour les *revenus des entreprises*, aucune donnée ne permet de mesurer l'*investissement* (ce deuxième résultat étant supposé augmenter avec le traitement). Les *actifs* sont en outre en baisse (effet significatif à 5 %). En Mongolie, trois résultats sont censés avoir des effets positifs. Cette conclusion vaut pour deux d'entre eux : *propriété des entreprises* et *consommation des ménages* (à 10 %). Cependant, bien que l'indice composite des *actifs* soit positivement impacté (à 10 %), l'effet n'est pas significatif (et même négatif) pour la valeur des *actifs*. Dans le cas du Maroc, où quatre résultats sont jugés positifs, nous renvoyons aux problèmes de fiabilité de cette RCT mentionnés ci-dessus. La synthèse du *Policy Bulletin* semble biaisée, ou au mieux très imprécise.

Compte tenu de ces lacunes, les coefficients élevés, mais non significatifs auraient été identiques même si la taille des échantillons s'était révélée suffisante. Ces résultats ont deux implications. Premièrement, ils remettent en question le postulat général selon lequel le microcrédit n'est pas « transformateur ». C'est peut-être vrai, mais finalement aucune donnée fiable ne permet de le prouver. Deuxièmement, DAHAL et FIALA (2020) concluent que « les études existantes [...] suggèrent que nous en savons encore très peu sur l'impact de la microfinance »¹⁵. Ce paradoxe, compte tenu des ressources consacrées aux RCT sur le microcrédit, est confirmé par MORDUCH (2020), l'un des meilleurs spécialistes au monde du microcrédit.

Une autre conclusion concernant la validité externe et interne concerne le fait qu'aucune des études de réplication (KINGI *et al.*, 2018 ; MEAGER, 2019 ; DAHAL et FIALA, 2020) n'ait pointé les erreurs que nous avons documentées dans notre réplication marocaine, notamment les plus évidentes, comme les déclarations des auteurs sur l'absence totale de contamination dans les groupes de contrôle, les comptages incohérents des ménages avant et après le *trimming*, et l'affirmation selon laquelle aucun *trimming* n'a été effectué dans l'enquête initiale. Cela ne fait que souligner les lacunes des réplifications « *push-button* » ou des réplifications qui appliquent des spécifications économétriques différentes aux mêmes données sans vérifier la fiabilité des données, des codages ou de l'échantillonnage d'origine.

15. Ce point est reconnu de façon détournée par les éditeurs du numéro spécial : « Les études individuelles peuvent manquer de preuves solides pour démontrer les effets transformateurs sur l'emprunteur moyen, mais elles manquent également de preuves solides contre les effets transformateurs » (BANERJEE *et al.*, 2015c : 3).

Tableau 5
Impact, références et publications.

	Bosnie- Herzégovine	Éthiopie	Inde	Mexique	Mongolie	Maroc
Impacts						
<u>Adhésion au crédit IMF</u>						
Source des données	Enquête	Enquête	Enquête	Admin., enquête	Enquête	Admin., enquête
Présence d'autres IMF	Oui	Oui	Oui	Oui	Oui	Oui
Effet de substitution/diminution d'autres formes de crédit	n. a.	Non	Oui (substitution)	Oui (diminution)	Oui (substitution)	Oui (substitution)
Impact	Positif (98,5 %)	Positif (25 %)	Positif (13 %)	Positif 8 % (enquête), 11 % (admin.)	Positif (50 %)	Positif 9 % (enquête), 17 % (admin.)
<u>Résultats (autres que l'adhésion au crédit)</u>						
Nombre	47	37	99	37	41	37
Nombre d'impacts significatifs (à 1 % ; 10 %)	0/47 (1 %) 3/47 (10 %)	0/37 (1 %) 5/37 (10 %)	1/99 (1 %) 13/99 (10 %)	3/37 (1 %) 9/37 (10 %)	0/41 (1 %) 10/41 (10 %)	6/37 (1 %) 17/37 (10 %)
Références, publications						
Nombre de références :	22	24	27	28	37	16
Dont RCT	5	11	11	20	10	8
Dont méthodologie/théorie	4	4	3	4	18	6
Dont autres méthodes de microcrédit	6	7	4	4	5	0
Autres	7	2	9	0	4	2
Nombre d'articles dans des revues académiques	1 (AEJ:AE)	2 (AEJ:AE ; Demography)	1 (AEJ:A)	1 (AEJ:AE)	1 (AEJ:AE)	1 (AEJ:AE)

Source : Florent Bédécarrats, Isabelle Guérin et François Roubaud, sur la base de BANERJEE *et al.* (2015c).

Notes : le nombre élevé de résultats dans la RCT indienne (99) est dû au fait que deux enquêtes finales ont été réalisées. Impact significatif (à 1 % ; 10 %) : nombre d'estimations d'impact associées à des valeurs p significatives au niveau de 1 % et au niveau de 10 %. IMF : institution de microfinance.

Résultats : des biais statistiques aux biais interprétatifs

La partie « Validité et portée du numéro spécial : évaluation critique » a exploré la fabrique des RCT sur le terrain et mis en exergue les nombreuses lacunes générées par des problèmes de validité interne et externe. Aux étapes de la collecte des données statistiques et de l'analyse économétrique fait suite l'étape de l'interprétation : « La beauté des évaluations randomisées est que les résultats sont ce qu'ils sont : nous comparons le résultat dans le traitement avec celui dans le groupe de contrôle, nous voyons s'ils sont différents et, si oui, de combien » (BANERJEE, 2007 : 115-16). Une analyse de la manière dont les *randomistas* transforment leurs données en déclarations scientifiques remet en question cette prétendue « beauté » des RCT.

Pris isolément, la plupart des résultats économétriques des six RCT n'ont aucun sens en eux-mêmes, sans parler de l'absence d'informations contextuelles. Les auteurs, notamment dans l'introduction générale, font cette interprétation dans un contexte très spécifique et au prix d'hypothèses implicites, mais puissantes, empruntées à une théorie comportementale du changement. Un cadre d'anthropologie et un cadre d'économie politique donneraient lieu à des conclusions très différentes. Notre but n'est pas de discréditer le processus d'interprétation, qui est inhérent à l'analyse des données, mais de démontrer que les *randomistas*, contrairement à ce qu'ils prétendent, ne peuvent y échapper. Les résultats ne sont pas « ce qu'ils sont », comme le montre également KABEER (2019) en employant des outils qualitatifs pour revisiter un terrain étudié par une RCT.

De plus, leur interprétation est fondée sur une « rhétorique de persuasion » (Labrousse, chap. 8, ce volume), qui consiste à faire table rase des recherches antérieures et à extrapoler (ici se pose à nouveau le problème de la validité externe), tout en faisant abstraction des questions spécifiques qui sont essentielles pour comprendre les impacts du microcrédit et que d'autres méthodes ont déjà abordées.

Faire table rase des recherches antérieures

Les résultats des *randomistas* sont souvent présentés comme des « découvertes » sans précédent, alors qu'ils ne sont souvent que la réplique de conclusions tirées d'études antérieures, notamment celles formulées à partir de méthodes non expérimentales qui ne sont presque jamais citées (LABROUSSE, 2010). L'introduction générale en est une bonne illustration. D'après la présentation qui en est faite, les résultats constituent les premières preuves scientifiques des impacts du microcrédit. « La base de preuves pour consacrer le microcrédit était assez mince » (BANERJEE *et al.*, 2015c : 1). Jusqu'à présent, les preuves empiriques disponibles étaient basées sur « des anecdotes, des statistiques descriptives ou des études d'impact qui ne permettent pas de distinguer la causalité de la corrélation » (*ibid.* : 1-2). Les auteurs prétendent prendre part aux « débats qui ont eu lieu dans les années 2000 et qui perdurent aujourd'hui » (*ibid.* : 2), mais,

de fait, ces débats se déroulent dans un cercle étonnamment cloisonné. Sur les 18 références mentionnées dans l'introduction générale, 12 (les deux tiers) proviennent des auteurs eux-mêmes et 17 (94,4 %) de membres du J-PAL. Un seul article échappe à ce principe endogamique.

Aucune étude non randomisée n'est citée. Pour ce qui est des six articles du numéro spécial, l'article sur le Maroc est tout aussi exclusif (seules des RCT sont mentionnées). Les autres le sont moins, mais de manière variable, comme le montre le tabl. 5. L'étude sur la Bosnie-Herzégovine est la plus pluraliste, avec un rapport RCT/non RCT de 0,8 ; ce rapport varie de 1,57 à 5 pour les autres.

Ce mépris des preuves disponibles non issues de RTC s'accompagne d'une tendance à extrapoler et à négliger les questions clés. Sans prétendre à l'exhaustivité, mais en se concentrant sur les points qui nous paraissent essentiels, nous abordons tour à tour les questions d'adhésion, de la création d'entreprises et de la liberté de choix, des transferts sociaux et de l'autosuffisance, ainsi que le problème du surendettement.

Adhésion

La faible adhésion constitue certainement le résultat le plus abouti du numéro spécial. De nombreux praticiens, décideurs et chercheurs s'échinent aujourd'hui encore à prédire un marché illimité, confondant ainsi exclusion financière et demande de crédit. Bien que ce résultat soit utile, sa véritable signification demeure limitée. Tout d'abord, il convient de noter que cet exercice n'a rien de nouveau. Certaines études mettent depuis longtemps en garde contre la faiblesse de la demande de microcrédit (JOHNSON et ROGALY, 1997 ; SERVET, 2006), notamment en fournissant des estimations quantitatives (KHANDKER *et al.*, 1998 ; HES et POLEDŇÁKOVÁ, 2013). En outre, les taux d'adhésion évoqués ici sont difficiles à comparer et à interpréter, étant donné la diversité des protocoles et des méthodes de randomisation (voir la section « Validité externe »). Il apparaît donc difficile d'évaluer la nature et la signification de la population cible et, par conséquent, de tirer des conclusions opérationnelles. De surcroît, les RCT ne disent rien sur les raisons de la faible adhésion : est-elle le reflet d'une demande intrinsèquement modeste et d'une faible propension à l'endettement et/ou d'une inadéquation de l'offre, étant entendu que les deux explications ne sont pas mutuellement exclusives ? Seules des données plus détaillées pourraient répondre à cette question, supposant une analyse fine des pratiques financières, comme celle des « agendas financiers » (*financial diaries*) (COLLINS *et al.*, 2009) et leurs implications sociales, morales et politiques (voir, par exemple, l'analyse qualitative du contexte marocain, négligé par les auteurs de la RCT marocaine : MORVANT-ROUX *et al.*, 2014).

Microcrédit, travail indépendant et liberté de choix

Les six études du numéro spécial tendent à s'accorder sur le fait que les impacts sur la création d'entreprises sont limités (significatifs dans deux cas seulement), l'expansion des entreprises existantes étant un impact plus fréquent (quatre cas).

Une amélioration de la rentabilité n'est constatée que dans un seul cas (Maroc), mais nous avons constaté plus haut que la validité interne de ces résultats laisse à désirer. En outre, même dans le cas d'une création ou d'une expansion d'entreprise, on n'observe aucun impact sur la croissance des revenus, soit parce que la rentabilité est faible, soit parce que les revenus du travail indépendant sont compensés par une baisse de revenus d'emplois salariés. Les auteurs de l'introduction générale prétendent ainsi tirer une conclusion inédite sur l'impact du microcrédit sur l'entrepreneuriat.

Cependant, depuis la fin des années 1980, de nombreuses études empiriques ont été menées pour mesurer l'impact du microcrédit¹⁶. L'étude systématique de DUVENDACK *et al.* (2011), conduite alors que les RCT commençaient à peine à voir le jour, tire deux conclusions. Premièrement, beaucoup d'études quantitatives, tant expérimentales (RCT comprises) qu'observationnelles, sont sujettes à de multiples biais¹⁷. Deuxièmement, lorsque les résultats sont valables, ils révèlent un impact limité et hétérogène, ce que Morduch a également observé à la fin des années 1990 dans son article pionnier sur les promesses partiellement non tenues du microcrédit (MORDUCH, 1999). Les résultats du numéro spécial n'ont donc rien d'inédit. Plus important encore, étant donné la complexité des chaînes causales induites par le microcrédit (DUVENDACK *et al.*, 2011) et l'hétérogénéité des effets et des types de microcrédit¹⁸, les RCT ne paraissent pas adaptées (BERNARD *et al.*, 2012). Enfin, la question des *randomistas* – le microcrédit fonctionne-t-il ou non ? – est mal posée. Ce que montrent les études rigoureuses (qu'elles soient quantitatives, qualitatives ou mixtes), c'est que certains types de microcrédit peuvent être utiles à certaines catégories de populations et dans certains contextes, mais pas d'autres (BÉDÉCARRATS, 2012 ; COPESTAKE *et al.*, 2016). Par exemple, les travaux de COPESTAKE *et al.* (2001 ; 2005) au Pérou et en Zambie et de BOUQUET *et al.* (2007) à Madagascar indiquent précisément quelles catégories de population bénéficient du microcrédit, et pourquoi, et, inversement, quelles catégories voient leur situation se dégrader, avec des propositions opérationnelles directes sur l'amélioration des services proposés afin d'en augmenter l'impact. Toujours à Madagascar et dix ans avant le numéro

16. BÉDÉCARRATS (2012) a identifié 154 études d'impact, contre 51 pour DUVENDACK *et al.* (2011).

17. Plusieurs répliquations d'études non expérimentales longtemps prises comme « preuves » de l'impact positif du microcrédit ont révélé de nombreux biais et une surestimation des impacts. Voir DUVENDACK et PALMER-JONES (2012) ; ROODMAN et MORDUCH (2014).

18. Nous donnerons l'exemple du microcrédit rural, qui est largement représenté dans le numéro spécial. Au-delà des modalités du crédit, quels sont les besoins en crédit (intrants, équipements, bétail, trésorerie pour financer la saison creuse, etc.), de quel type d'agriculture parlons-nous (cultures commerciales ou vivrières, agriculture en zone sèche ou pluviale, intensive ou extensive, familiale ou professionnelle, indépendante ou contractuelle par intégration dans des secteurs agro-industriels ou des coopératives de producteurs, etc.) et quelle est la nature des économies rurales (degré de monétarisation, éloignement et qualité des infrastructures, opportunités de revenus non agricoles) ? Et, surtout, de quel type d'IMF parlons-nous ? Le statut (but lucratif/non lucratif) est certes un facteur de distinction (spécifié dans le numéro spécial), mais d'autres questions essentielles se posent, notamment le mode de gouvernance, le degré d'intégration et d'adaptation aux réalités locales et la capacité à concevoir des produits adaptés à la demande locale. Compte tenu de cette diversité, il est absurde de parler de « microcrédit rural ». Sur cette diversité, voir par exemple MORVANT-ROUX (2009).

spécial, notre propre évaluation de l'impact d'une IMF locale basée sur une approche quasi expérimentale suggère trois grandes caractéristiques stylisées, présentées plus tard comme des « découvertes » par les *randomistas*. L'impact du microcrédit n'est pas « transformateur », les impacts sont hétérogènes au regard de la distribution des entreprises par taille et le contexte importe : le microcrédit est plus bénéfique en période de croissance qu'en période de crise (GUBERT et ROUBAUD, 2011).

Pour comprendre l'hétérogénéité des impacts (et en tirer des conclusions opérationnelles), une conception distincte des mécanismes de causalité est nécessaire : le but n'est pas de « faire la différence », mais de penser en termes de « mécanisme » et de « processus » (SHAFFER, 2015). De plus, étant donné les nombreuses externalités, il apparaît tout aussi restrictif de se concentrer sur l'impact individuel. Très peu d'études ont appliqué des modèles d'équilibre général au cas du microcrédit à méso-échelle (pour une exception, voir MAHJABEEN, 2008). Les examens des externalités ont été principalement réalisés par des analyses d'économie politique, puisque c'est précisément l'analyse de l'ancrage des IMF dans leur environnement social, culturel, politique et économique et des externalités qui a un puissant effet sur l'adhésion des produits et, partant, sur leur impact (COPESTAKE *et al.*, 2016). Des études d'impact convaincantes et utiles conduites dans des zones rurales ont démontré le rôle clé des innovations financières ancrées dans les territoires locaux – capables de développer des produits spécifiques conçus à l'échelle locale (crédits-relais, fonds de garantie et crédit-bail) et associés à d'autres mesures (contrats de métayage, entrepôt à récoltes, assistance technique, etc.) – pour inciter les petits agriculteurs à participer à différentes chaînes de valeur (BOUQUET *et al.*, 2007 ; BASTIAENSEN et MARCHETTI, 2011), tout en rencontrant souvent des effets de seuil (DOLIGEZ, 2002). Les effets sont parfois discutables, comme lorsque le microcrédit accélère les processus de migration, puisque la migration devient nécessaire pour rembourser les microcrédits (MORVANT-ROUX, 2013 ; BYLANDER, 2014). Leur nature est parfois plus politique et culturelle qu'économique. En Égypte, par exemple, l'introduction du microcrédit perturbe les valeurs locales – définies au sens large comme ce qui fait sens pour les gens – et, par conséquent, les processus de reconnaissance, d'identité et de socialisation (ELYACHAR, 2006). Dans l'Inde rurale du Sud, la présence massive d'IMF dans certains territoires reconfigure les rapports de force locaux et les chaînes de patronage en les féminisant (GUÉRIN et KUMAR, 2017). Ces résultats (et les questions qui s'y rapportent) sont très éloignés de ceux des *randomistas*. Et pourtant, si nous voulons vraiment comprendre ce que le microcrédit change dans la vie des gens, c'est précisément ce type de questions larges qu'il faut nous poser.

Outre ces études approfondies, qui reposent systématiquement sur une bonne connaissance des contextes locaux dans le temps, il est utile de mentionner d'autres méthodes plus légères, conçues pour identifier rapidement les caractéristiques des clients (et des non-clients) et leur mode d'utilisation des services et en tirer des recommandations en vue d'améliorer la qualité de l'offre, qui

demeure la principale question récurrente posée par les fournisseurs de micro-crédit¹⁹.

Revenons maintenant au numéro spécial. Non seulement les auteurs n'ajoutent rien de fondamentalement nouveau aux débats, mais leur interprétation des résultats quantitatifs pose problème. La micro-entreprise peut refléter l'absence de choix, et non la multiplication des choix. Une grande partie des micro-entrepreneurs, condamnés au travail indépendant faute d'emploi rémunéré, relève plus de l'auto-exploitation analysée par CHAVANOV (1966 [1925]) que de l'entrepreneuriat schumpétérien (LAUTIER, 2004). Le cas de la Mongolie est instructif à cet égard. La RCT montre que l'accès au crédit solidaire permet aux femmes de créer de nouvelles micro-entreprises, mais pour des revenus négatifs, alors que leur temps de travail augmente de plus d'un tiers (sans variation du temps passé au foyer). Ces effets négatifs sont principalement observés chez les femmes moins instruites (ATTANASIO *et al.*, 2015 : 105, note 21). Les auteurs pensent que la rentabilité peut s'améliorer une fois le crédit remboursé (ATTANASIO *et al.*, 2015 : 115). On retrouve ici le problème de la temporalité qui, comme nous l'avons déjà souligné, limite fortement l'intérêt des méthodes randomisées (LABROUSSE, 2010 ; BÉDÉCARRATS *et al.*, 2019b). Ces femmes peuvent en effet avoir choisi de se lancer dans l'aventure entrepreneuriale, ce qui peut expliquer l'amélioration de la consommation (les résultats indiquent une consommation plus importante et plus saine). Mais que signifie ce « choix » et, surtout, quelles en sont les conséquences s'il entraîne un accroissement des responsabilités et éventuellement un désengagement des autres membres du ménage (et donc des inégalités intrafamiliales) ? Les données quantitatives ne permettent pas de tirer une conclusion à ce sujet et les auteurs de la RCT ne portent pas de jugement particulier. Une interprétation robuste exigerait d'autres types de données, quantitatives ou qualitatives. Les auteurs de l'introduction générale, pour leur part, se concentrent uniquement sur la dimension de « liberté de choix », sans mentionner les effets potentiellement négatifs de ces « choix » sur les femmes, en particulier les plus défavorisées.

Microcrédit, dépenses sociales, transferts sociaux et autosuffisance

Si les effets en termes d'activité et de revenus ne sont pas concluants, les auteurs de l'introduction générale observent ce qu'ils décrivent comme des effets positifs sur deux indicateurs : les « dépenses non essentielles », signe d'une discipline accrue et de meilleures compétences de gestion, et la réduction des « transferts sociaux », signe d'une plus grande autonomie. Les « dépenses non essentielles » incluent les « biens de tentation » et ont diminué dans quatre pays (elles n'ont pas été mesurées en Éthiopie et les résultats n'étaient pas significatifs en Mongolie) :

19. Les outils élaborés par AIMS (Assessing the Impact of Microenterprise Services) et Imp-act, qui ont été dénigrés pour leur manque de méthode quantitative sophistiquée, en sont des exemples. Ces outils peuvent difficilement « prouver » l'impact à grande échelle, mais ils se sont révélés très utiles pour « améliorer » et diversifier l'offre de services de microfinance.

alcool et cigarettes en Bosnie-Herzégovine, cigarettes, sucreries et sodas au Mexique, alcool, tabac, feuilles de bétel, jeux de hasard et aliments consommés hors du foyer en Inde. Ces dépenses incluent également les festivités, avec des baisses observées en Inde et au Maroc.

Les auteurs avancent plusieurs explications de cette réduction des « biens de tentation » : contraintes de remboursement et d'investissement, meilleure auto-discipline et plus grande participation des femmes à la prise de décision. Le déclin des dépenses de tentation constitue l'un des principaux résultats de la RCT indienne présentés dans le résumé. Les auteurs de l'étude prennent soin de préciser que ce sont les populations elles-mêmes qui décrivent ces biens comme des « biens de tentation », en ce sens qu'elles souhaiteraient réduire ce type de consommation (BANERJEE *et al.*, 2015b : 24). Mais le fait que les gens expriment cette préférence (une observation aux origines vagues, qui semble plutôt relever de l'« anecdote », dont l'utilisation est soulignée par Labrousse, chap. 8, ce volume), pourrait bien indiquer qu'ils ont intégré les discours moralisateurs fréquemment assénés par les organisations de développement (IMF comprises), et ce, depuis la période coloniale²⁰.

Au-delà de la dimension moralisatrice des conclusions des *randomistas*²¹, une analyse détaillée de la signification et du rôle de ces dépenses pourrait apporter un éclairage différent. Sur le sujet de l'alcool, personne ne conteste que la consommation excessive pose un problème de santé publique. Pourtant, si nous voulons vraiment comprendre ce type de consommation et concevoir des lignes d'action, il est essentiel de reconnaître la dimension sociale et politique de l'alcool. Comme beaucoup d'autres biens de tentation, et contrairement à ce que suggère l'économie comportementale, il ne s'agit pas d'un bien qui se définit uniquement par son « utilité immédiate » (BANERJEE et MULLAINATHAN, 2010). L'alcool peut jouer un rôle social en ce sens qu'il permet aux travailleurs d'endurer un travail très exigeant physiquement et d'accéder à des espaces de socialisation, et donc à des informations stratégiques (les bars constituent souvent des lieux privilégiés pour négocier des contrats de travail et des commandes ; PICHÉRI, 2018). L'alcool peut jouer un rôle politique lorsqu'il donne aux travailleurs l'occasion de formuler vis-à-vis des employeurs et des patrons des revendications qui sont plus facilement acceptables sous l'effet de l'ébriété (SCOTT, 1977). Et, surtout, l'alcool est souvent proposé délibérément par les employeurs et les recruteurs de main-d'œuvre afin de fidéliser les travailleurs (PICHÉRI, 2018). Il est donc fallacieux de dire que le sacrifice ou une plus grande maîtrise de soi suffiraient pour lutter contre ces « tentations ».

20. En Inde, par exemple, les colons britanniques et les missions chrétiennes du début du XIX^e siècle dénonçaient déjà l'imprévoyance et la prodigalité des pauvres (CEDERLÖF, 1997 ; HARDIMAN, 2000).

21. Les discours des *randomistas* rappellent la morale victorienne de la révolution industrielle européenne, légitimée par les arguments des économistes néoclassiques de l'époque. Face à l'extrême pauvreté du monde ouvrier pendant la révolution industrielle britannique, certains ont déploré le manque d'autosuffisance des pauvres, leur manque de prévoyance et le gaspillage de leurs revenus en alcool, et ont plaidé pour des cours d'éducation financière plutôt que pour des hausses de salaire (voir par exemple JEVONS, 1883 : 196-200, 205).

De même, les frais de restauration (repas et thé) hors du domicile ne sont pas uniquement des opportunités manquées « d'épargne lucratives » (BANERJEE et DUFLO, 2011 : 170). Les restaurants de rue et les gargotes de thé sont des lieux éminemment stratégiques. Dans une économie informelle opaque, structurée par des relations interpersonnelles, ces espaces permettent aux commerçants de se tenir mutuellement informés de la situation du marché, de l'évolution des prix, des opportunités à saisir, des sources de financement possibles, des risques de contrôles fiscaux ou policiers, etc. Les petits entrepreneurs y cultivent des liens d'échange et de soutien mutuel, qui ont une fonction souvent décisive pour la survie de leur entreprise.

En ce qui concerne les dépenses consacrées aux rituels sociaux et religieux, l'anthropologie a depuis longtemps démontré que la « richesse sociale » constitue un facteur essentiel de réussite et de protection (GUYER, 1997) et qu'« investir » dans les relations sociales peut, dans certaines situations, se révéler beaucoup plus rationnel que d'essayer d'économiser de l'argent en se coupant de son entourage (NAROTZKY et BESNIER, 2014). Au-delà des *randomistas*, la question des « taxes communautaires » et de leur rapport coût-bénéfice en termes de protection a fait l'objet de diverses études par des économistes du développement. Mais ces études tiennent rarement compte de la complexité des circuits financiers auxquels ces dépenses donnent lieu et de leur caractère durable. Une analyse conduite en Inde sur la corrélation entre les dépenses festives et les invitations à déjeuner montre, par exemple, que ces dépenses font office de filet de sécurité (RAO, 2001). De plus, ce que les économistes considèrent comme une dépense est parfois vu comme une créance ou une épargne, puisqu'il s'ensuivra une compensation. Toujours en Inde, la comptabilisation de l'ensemble des dettes et des créances générées au fil du temps par les dépenses cérémonielles, dont les familles ont bien conscience puisqu'elles les calculent en ces termes, montre que la richesse financière nette des familles est radicalement différente de celle que suggère une analyse en termes de « dépenses » (GUÉRIN *et al.*, 2019). Cela contredit l'idée avancée par les *randomistas* à propos des biais de court terme dont souffriraient les pauvres (BANERJEE et DUFLO, 2011 : 183-204).

En ce qui concerne les transferts sociaux, sur les huit estimations retenues (qui portent sur des transferts en provenance de la famille ou l'État), cinq sont négatives. Cette observation conduit les éditeurs du numéro spécial à conclure que l'« autosuffisance » s'est améliorée, un facteur qui est jugé de manière positive²². Cette interprétation est à la fois risquée – il n'y a aucune raison de croire que la baisse des transferts depuis la famille et les amis soit jugée positive ou considérée comme une source de bien-être par les personnes concernées – et normative, à l'instar des interprétations précédentes. Là encore, l'anthropologie est d'une aide précieuse pour élucider le rôle décisif des interdépendances

22. Il convient toutefois de noter que cette interprétation est celle des auteurs de l'introduction, et non des auteurs des articles, qui ne commentent pas ce résultat, mais en soulignent bien l'ambiguïté. Pour ce qui est de la Mongolie, les auteurs mentionnent, par exemple, que « le renforcement de la discipline financière au sein du groupe peut avoir pour effet de perturber les systèmes informels de crédit et d'assurance basés sur la parenté et d'autres liens sociaux » (ATTANASIO *et al.*, 2015 : 114).

sociales, tant en termes de protection matérielle que d'identité. Au-delà des *randomistas*, de nombreux acteurs du monde du développement – décideurs politiques, praticiens et certains chercheurs –, considèrent la dépendance à la fois comme un problème politique (l'aide est coûteuse) et comme un problème moral (la dépendance est réputée être incompatible avec la liberté individuelle). Cependant, dans de nombreux contextes, les liens et la dépendance vis-à-vis des autres constituent à la fois un mode d'action et une stratégie délibérée. Par conséquent, l'agentivité des gens et leur « liberté de choix » se traduisent plutôt par la capacité à choisir certaines formes de dépendance et d'interdépendance²³.

En définitive, la conclusion de l'introduction générale sur l'amélioration de l'autosuffisance, ainsi que celle sur la « liberté de choix », est induite par des interprétations arbitraires des résultats économétriques (voire des extrapolations des conclusions de certaines des RCT). Ces interprétations sont sous-tendues par une conception singulière de l'autonomie et de la liberté individuelles, et donc par leur propre théorie du changement, qui voit les gens comme des atomes isolés, faisant fi des multiples rôles que jouent les interdépendances sociales à différents niveaux et jugeant implicitement néfastes ces interdépendances. Ces deux conclusions – « autosuffisance » et « liberté de choix » – ont néanmoins été incluses dans le *Policy Bulletin* (J-PAL et IPA, 2015), qui a ensuite été largement diffusé par de nombreux blogs et réseaux de discussion et considéré comme un résultat incontestable de cette recherche.

Microcrédit et surendettement

Une conclusion majeure du numéro spécial est que le microcrédit n'est pas le « piège de la dette » dénoncé par ses détracteurs. Tout d'abord, il convient de noter qu'aucune étude scientifique n'est mentionnée dans l'introduction générale, comme si le « piège de la dette » était une réalité anecdotique sans aucun fondement empirique. Il est vrai que les médias ont abondamment parlé des diverses crises de remboursement des microcrédits (tout comme ils avaient fait l'éloge du microcrédit à ses débuts). Cependant, presse mise à part, il existe un vaste corpus scientifique traitant du surendettement des ménages dans les pays du Sud et du rôle joué par le microcrédit (SCHICKS et ROSENBERG, 2011 ; GUÉRIN *et al.*, 2013b ; 2015 ; SCHICKS, 2013), y compris dans les pays couverts par le numéro spécial. Un certain nombre de problèmes se posent ici.

Le premier concerne la validité externe, où l'extrapolation se fait sans tenir compte de la singularité des contextes étudiés, ni de la dimension « marginale » des zones étudiées (WYDICK, 2016). Les six RCT se sont concentrées sur des zones et des populations qui étaient censées ne pas avoir accès au microcrédit²⁴.

23. Pour avoir un aperçu général de la façon dont l'anthropologie aborde cette question, voir par exemple FERGUSON (2015).

24. Comme mentionné ci-dessus, cette « virginité » était en fait un leurre et toutes les populations témoins avaient accès au microcrédit. Toutefois, le marché n'était pas aussi saturé qu'il aurait pu l'être ailleurs, de sorte que le risque de surendettement était moindre.

Or, par définition, le problème du surendettement y est moins aigu que dans les zones et les populations précédemment exposées au microcrédit. C'est donc une tautologie de dire qu'il n'y a pas là de « piège de la dette ». Le surendettement de certains clients du microcrédit a pourtant été documenté et parfois mesuré dans quatre des pays étudiés²⁵. Le fait que les RCT ne l'aient pas quantifié ne leur permet pas de conclure que le piège de la dette n'existe pas. Contrairement à ce que suggèrent les auteurs de l'introduction générale, la littérature disponible ne se contente pas d'« anecdotes ». Les chercheurs ayant travaillé sur ce sujet démontrent (le plus souvent de manière qualitative) le rôle du microcrédit en se basant sur une analyse détaillée de ses caractéristiques spécifiques par rapport à d'autres sources d'endettement, en particulier la rigidité des conditions de remboursement et la faible tolérance aux impayés. Dans certains contextes et certaines IMF, cette tolérance zéro prend la forme de procédures coercitives de remboursement²⁶. Ces chercheurs proposent également une analyse nuancée et contextualisée, mettant en évidence le rôle de l'économie globale dans le contexte mondial (notamment la stagnation et la baisse des revenus réels face à des besoins croissants), ainsi que le rôle ambivalent du microcrédit (pour certains emprunteurs, le microcrédit peut être un moyen de rembourser des dettes informelles et de réduire le surendettement)²⁷. Le lien de causalité entre le microcrédit et le surendettement peut ne concerner qu'une minorité de clients du microcrédit (ce qui nous ramène à la question de l'hétérogénéité), mais ses répercussions (appauvrissement, exclusion sociale, suicide, etc.) (SCHICKS, 2013) sont suffisamment tragiques pour justifier que les *randomistas* prennent le phénomène plus au sérieux.

Le second problème est l'extrapolation des six études de cas par les auteurs de l'introduction. Même dans les zones et les populations récemment exposées au microcrédit, le surendettement ne peut être exclu. La RCT de Bosnie-Herzégovine a été conduite dans un contexte de crise de surendettement avéré, que les auteurs mentionnent comme élément de contexte (AUGSBURG *et al.*, 2015 : 185). Cette RCT conclut spécifiquement que le groupe de traitement avait des difficultés de remboursement (AUGSBURG *et al.*, 2015 : 199-201) et que celles-ci constituent un symptôme potentiel de surendettement²⁸. La RCT ne permet de conclure ni à l'existence du surendettement ni au rôle du microcrédit. Cela étant, l'existence d'un « piège de la dette » ne peut être exclue.

25. Pour le Mexique, voir MORVANT-ROUX (2013), ANGULO SALAZAR (2013), HUMMEL (2013), ROZAS (2014). Pour l'Inde, voir TAYLOR (2011), GUÉRIN *et al.* (2013b), JOSEPH (2013), PRATHAP et KHAITAN (2016). Pour la Bosnie-Herzégovine, voir BATEMAN (2010), MAURER et PYTKOWSKA (2011), OPEM et GORONJA (2013). Pour la Mongolie, voir JAVOY et ROZAS (2013).

26. En Inde, par exemple, la poursuite des mauvais payeurs sur leur lieu de travail ou à leur domicile, les dénonciations et insultes publiques, la sollicitation de parents, les menaces physiques, la confiscation de biens et de documents administratifs ; dans certains cas, les plus récalcitrants ont été ligotés sur la place publique ou en plein soleil (ARUNACHALAM, 2011 ; SERVET, 2011).

27. À l'heure où nous finalisons ce chapitre (octobre 2019), les Nations unies viennent de se saisir de cette question en commandant un rapport sur le sujet, ce qui semble indiquer que le problème existe bel et bien : <https://www.ohchr.org/EN/Issues/Development/IEDebt/Pages/ReportPrivateDebt.aspx>.

28. Les défauts de paiement peuvent aussi être des défaillances « stratégiques » exprimant un refus de rembourser, notamment dans le contexte d'une crise d'impayés.

Dans la RCT mongole, les auteurs prennent soin de préciser que leur étude ne mesure pas le surendettement, mais uniquement les défauts de remboursement, qui sont deux choses distinctes²⁹. L'introduction du numéro spécial ne fait pas référence à ces précisions.

Au Maroc, une étude qualitative menée par l'un d'entre nous en même temps que la RCT a conclu à une faible propension à l'endettement dans les zones rurales, pour des raisons culturelles et religieuses (MORVANT-ROUX *et al.*, 2014). Ce constat général, valable « en moyenne », n'exclut cependant pas les problèmes de surendettement dans une fraction de la population. Étant donné que le Maroc a également connu une crise de remboursement du microcrédit (que les auteurs ne mentionnent pas, bien qu'elle ait eu lieu pendant la RCT), les IMF concentrent leur offre sur une minorité de clients jugés solvables et fiables. Ces clients sont donc surexposés au microcrédit, et certains d'entre eux rencontrent effectivement des problèmes de surendettement (MORVANT-ROUX et ROESCH, 2015).

Tout comme la Bosnie-Herzégovine, l'Inde a été frappée par d'importantes crises d'impayés dans le domaine du microcrédit : dans le district de Krishna, dans l'Andhra Pradesh, en 2006, puis dans une petite ville du Karnataka en 2009 et dans tout l'État de l'Andhra Pradesh en 2010. Les analyses de cette crise, tant quantitatives que qualitatives, ont révélé l'existence d'un problème de surendettement pour une partie des clients. Le surendettement des populations pauvres, avec ou sans microcrédit, a également été documenté en dehors des zones de crise de la dette, y compris dans les zones urbaines. Comme indiqué plus haut, la RCT indienne a été menée de 2005 à 2010 dans des quartiers périphériques d'Hyderabad nouvellement exposés au microcrédit. Mais comment est-il possible d'extrapoler à partir de cette étude de cas très spécifique alors qu'il existe un vaste ensemble de preuves démontrant l'existence du surendettement ? Sur cette question, l'article de (BANERJEE *et al.*, 2015b : 23) ne cite qu'un seul article de presse : « Les anecdotes sur les entrepreneurs très prospères ou les emprunteurs très endettés ne nous disent rien sur l'effet de la microfinance sur l'emprunteur moyen, et encore moins sur l'effet de l'accès à la microfinance pour le ménage moyen ». Au vu de l'état d'alerte sur le niveau de surendettement des populations indiennes pauvres, et compte tenu de l'extrême spécificité des zones qu'ils étudient, n'est-ce pas leur propre étude qu'il faudrait peut-être qualifier d'anecdotique ?

Enfin, on peut se demander si la mesure de l'endettement des ménages a été correctement effectuée. La collecte de données fiables sur l'endettement exige un certain nombre de précautions pour les raisons suivantes : le tabou de l'endettement, exacerbé lorsque les IMF prétendent éradiquer l'emprunt informel au motif qu'il encourage les clients à dissimuler leurs dettes informelles, la diversité des terminologies employées et l'éventail des dettes susceptibles d'être détenues par différents membres de la famille sans qu'ils ne partagent nécessairement cette information. Compte tenu des approximations observées aux autres étapes de

29. Comme certaines défaillances peuvent être stratégiques, de bons taux de remboursement peuvent masquer les sacrifices consentis pour honorer des dettes, ce que reconnaissent les auteurs de la RCT en Mongolie (ATTANASIO *et al.*, 2015 : 114, note 25).

la collecte et de l'analyse des données (partie « Validité et portée du numéro spécial : évaluation critique »), il n'est pas déraisonnable de mettre en doute la capacité des *randomistas* à concevoir un questionnaire rendant dûment compte de l'endettement des ménages. Il convient toutefois de noter que cette difficulté n'est pas propre aux *randomistas*. La collecte de données fiables sur les revenus dans les pays du Sud a nécessité des décennies d'apprentissage pour adapter les outils statistiques aux contextes où les ménages jonglent avec différentes sources de revenus, y compris des sources informelles. Ce travail reste encore à faire sur la dette, qui demeure mal mesurée et souvent sous-estimée.

Conclusion et discussion

Compte tenu des nombreuses limites et lacunes que nous avons constatées avec la randomisation, appliquée ici au microcrédit, on peut se demander pourquoi les RCT ont connu un tel succès dans les milieux académiques, médiatiques et politiques. Nous avons déjà analysé les raisons de cette contradiction (BÉDÉCARRATS *et al.*, 2019b) en explorant l'économie politique de ce qui est devenu une véritable industrie (Ravallion, chap. 1, ce volume). Comme pour toute industrie, le marché de l'évaluation d'impact met en lien une offre et une demande. Nous avons exploré ces deux éléments en détail, en montrant que la demande est double, stimulée à la fois par la communauté des bailleurs de fonds et par le monde académique, tandis que l'offre est largement façonnée par des entrepreneurs scientifiques qui semblent avoir créé un nouveau « *business model* » en vue d'imposer un monopole et une position de rente sur le marché de l'impact de l'évaluation. La manière dont les données ont été produites et analysées, comme nous l'avons fait ici, est une illustration de cette stratégie de domination. Outre la technique de la table rase (partie « Faire table rase des recherches antérieures »), trois autres stratégies paraissent essentielles : s'affranchir d'une « culture de la donnée », ignorer la critique et contourner certaines règles de l'éthique scientifique.

S'affranchir d'une « culture de la donnée »

Les nombreuses erreurs de collecte et de saisie des données observées dans la RCT marocaine semblent suggérer un certain manque d'expérience et de connaissances, comme si les compétences purement techniques requises pour la deuxième étape (économétrie : traitement des questions de biais, sélection et identification d'un contrefactuel) dispensaient les chercheurs de l'obligation de se doter du savoir-faire requis pour la première étape (collecte de données de bonne qualité). Dans quelle mesure cette considération s'applique-t-elle aux autres RCT ? Malheureusement, cette question reste pour l'heure ouverte, car seules des répliques complètes pourraient y répondre. Cela étant, il est clair

que les *randomistas* ont tendance à ignorer les débats autour de la collecte de données (comme ils le font pour la question de l'éthique : voir Abramowicz et Szafarz, chap. 10, ce volume). Dans la plupart des protocoles de recherche empirique quantitative, il y a une véritable division du travail entre les collecteurs et les analystes de données : les premiers sont des statisticiens, les seconds sont des économistes (économétriciens ou spécialistes de thématiques spécifiques). À quelques exceptions près (DEATON, 1997 ; GROSH et GLEWWE, 2000), peu de personnes peuvent se targuer de couvrir les deux extrémités du spectre. Il s'agit de métiers à part entière, qui requièrent des compétences et une formation distinctes. Les statisticiens sont responsables de l'exactitude de la mesure, les économistes de sa pertinence, de son analyse et des relations et interactions entre les données. Ces deux activités sont indispensables pour générer des résultats « raisonnables », même si les statisticiens ont moins de prestige social que les économistes (DESROSIÈRES, 2013a). Compte tenu des compétences impliquées et du mode de fonctionnement des revues académiques, tous les efforts sont concentrés en amont sur la conception d'un processus de randomisation « intelligent » et en aval sur les estimations économétriques des impacts, le but étant de publier des articles dans des revues de premier plan.

La déconnexion entre les chercheurs et le terrain est une autre illustration de ce déni à l'égard de la culture des données. Cette déconnexion est particulièrement aiguë chez J-PAL. Son organisation hiérarchique impose une stricte division du travail entre les chefs de projet, les doctorants et les personnels de terrain (superviseurs et enquêteurs). Ces derniers se voient confier des responsabilités considérables pour lesquelles ils ne sont sans doute pas suffisamment formés (JATTEAU, 2018). Cette division du travail est une pratique fréquente dans le domaine des sciences naturelles et des sciences du vivant, mais elle n'empêche pas les chefs d'équipe de rester en contact régulier avec la chaîne de production des données, y compris pour les expériences *in vivo*. De plus, les équipes sont tenues de respecter des protocoles précis pour valider la rigueur des expérimentations réalisées. Ce n'est probablement pas le cas ici, étant donné les dizaines de RCT impliquant les personnalités les plus éminentes de la mouvance RCT (BÉDÉCARRATS *et al.*, 2019b). Cette déconnexion a été exacerbée par l'expansion exceptionnellement rapide de J-PAL, comme indiqué plus haut.

Cette croissance, conjuguée à une gouvernance extrêmement centralisée, fait que seule une poignée de chercheurs dirigent un nombre considérable d'expérimentations, ce qui fait planer le doute sur leur capacité réelle à s'impliquer sérieusement sur chaque RCT (et fait le lit de la déconnexion avec le terrain). En février 2019, Esther Duflo avait ainsi 64 RCT à son actif, soit un peu plus de quatre nouvelles RCT par an. Mais c'est Dean Karlan qui est de loin le plus prolifique, avec 100 RCT (dont 42 en cours). Dès lors, on est en droit de se demander ce qu'ils donnent réellement de leur personne pour produire chacun des résultats des RCT qu'ils signent. De fait, la signature d'un chercheur *randomista* de haute volée semble plus avoir valeur de laissez-passer pour se faire publier dans une revue de premier plan, dans le cadre d'une stratégie globale de randomisation, que de gage de qualité de la recherche.

Ignorer les critiques

Si les *randomistas* ont élaboré un discours universel sur l'impact du micro-crédit à l'issue de ce numéro spécial (et de publications ultérieures), d'autres acteurs ont pour leur part tiré des conclusions différentes de ces mêmes études (KABEER, 2019). Là encore, la RCT marocaine en est une illustration typique. Dès 2009, alors que l'enquête finale était encore en cours, le bailleur de la RCT a commencé à partager publiquement son point de vue sur la méthode RCT en se basant sur la RCT marocaine et sur une autre étude RCT menée simultanément au Cambodge. Les conclusions étaient claires : elles mettaient en évidence les difficultés rencontrées par la méthode pour produire des évaluations d'impact rigoureuses en raison des multiples manquements au protocole que l'équipe de recherche du bailleur avait partiellement identifiés (problème de représentativité et changement de produit) et des contraintes de temps qui obligeaient à se focaliser sur des impacts de court terme. Si les conclusions de l'équipe de recherche du bailleur de fonds ont été présentées publiquement et publiées à de nombreuses reprises (BERNARD *et al.*, 2012), elles sont restées lettre morte pour l'équipe de la RCT (BÉDÉCARRATS *et al.*, 2019b).

Notre propre expérience de la RCT marocaine, bien qu'illustrative, donne un bon exemple de ce que pourrait être une stratégie consistant à ignorer les critiques, jusqu'à un certain point. Au cours de nos recherches critiques dans les champs du développement, nous avons invité certains des partisans les plus virulents des RCT à engager un débat scientifique (une controverse) en de nombreuses occasions (sessions dédiées lors de conférences internationales). Nous n'avons à ce jour reçu aucune réponse. Nous avons également invité dix des plus célèbres *randomistas* à participer à cet ouvrage collectif afin d'équilibrer les points de vue sur les RCT. Ils ont tous refusé. Concernant notre examen critique de la RCT marocaine, nous avons informé les auteurs du déroulement, puis de la publication de notre réplique (BÉDÉCARRATS *et al.*, 2019a). En même temps, nous avons rédigé un *Comment* et suggéré que l'*AEJ:AE* le publie avec une *Answer to the Comment* des auteurs, comme le veut la pratique dans de nombreuses revues. L'*AEJ:AE* a refusé l'offre au motif que la revue ne publie pas de commentaires. Enfin, lorsque notre article a été évoqué dans des blogs influents et dans la presse, CRÉPON *et al.* (2019) ont rédigé un *Rejoinder* (51 pages) en recourant à des analyses sophistiquées pour expliquer que leurs résultats originaux étaient robustes : *double post lasso procedure*, *Benjamini-Hochberg false discovery rate correction of multiple testing*, *bayesian hierarchical model*, et analyse d'apprentissage automatique, entre autres, concluant que notre réplique n'était pas scientifique. Ils ont publié leur *Rejoinder* sur leur site internet et nous ont enjoint de le publier sur le site DIAL, ce que nous avons fait en bonne et due forme. Ils ont également informé la hiérarchie de l'AFD. L'*International Journal for Re-Views in Empirical Economics* (IREE) a suggéré que les deux parties publient une version courte du *Rejoinder* avec notre réponse (*Rebuttal of the Rebuttal* ; BÉDÉCARRATS *et al.* [2019c]). Au vu des conclusions totalement contradictoires des deux documents, nous avons proposé de demander l'avis

d'un tiers pour déterminer si nous devions retirer notre réplique (BÉDÉCARRATS *et al.*, 2019a) ou le papier initial (CRÉPON *et al.*, 2015) en fonction de la conclusion. Une fois de plus, ils ont décliné l'invitation. Ces épisodes illustrent deux caractéristiques de la stratégie *randomista*. Premièrement, contrairement à l'un des principaux arguments de vente des RCT (la simplicité de la méthode, par rapport à la « boîte noire » des méthodes économétriques alternatives), ce type de RCT est extraordinairement complexe. Dans leur *Rejoinder*, ils ont rajouté de la complexité à un protocole de randomisation déjà extraordinairement complexe (l'un des trois paradoxes que nous avons cherché à expliquer dans BÉDÉCARRATS *et al.* [2019b]). Deuxièmement, ils ont contourné les normes scientifiques en ne mentionnant pas leurs codes, en refusant une révision par les pairs de leur *Rejoinder* et, enfin, en éludant une controverse scientifique équilibrée.

Contourner l'éthique scientifique

En plus de faire fi de tout ce qui ne relève pas des RCT, les *randomistas* ont éludé certaines règles de base de la conduite scientifique. Ce problème semble gagner de l'ampleur dans l'ensemble de la communauté scientifique (HECKMAN et MOKTAN, 2018). Or, s'il n'est certes pas spécifique au J-PAL ou à la communauté des *randomistas*, il est particulièrement patent ici. Dans le monde de la recherche, la validation des connaissances est basée sur le principe de l'« examen par les pairs », c'est-à-dire un examen collectif pratiqué par des chercheurs qui jugent de manière critique et anonyme le travail de leurs homologues. Mais, pour ce faire, de nombreuses règles éthiques doivent être respectées, à commencer par la gestion des conflits d'intérêts entre les auteurs et les membres des comités de rédaction des revues. Le favoritisme éditorial est un processus reconnu et démontré, notamment chez les économistes (FOURCADE *et al.*, 2015). Le numéro spécial est révélateur à cet égard. Les trois éditeurs scientifiques du numéro sont tous membres du J-PAL (BANERJEE *et al.*, 2015c). Outre l'introduction générale, chaque éditeur a cosigné un article et deux d'entre eux étaient membres du comité de rédaction (Banerjee et Karlan). Esther Duflo est à la fois rédactrice en chef (et fondatrice) de la revue et co-auteure de deux des six articles. De surcroît, comme près de la moitié des auteurs des articles (11 sur 25) sont également membres du J-PAL et comme quatre autres sont des professeurs ou des doctorants affiliés au J-PAL, la revue a quelque peu dérogé aux principes d'examen par les pairs censés régir toute publication scientifique. Ce seul exemple illustre l'extraordinaire densité des liens entre les promoteurs des RCT, densité par ailleurs démontrée à une échelle beaucoup plus large par JATTEAU (2016).

Que reste-t-il du numéro spécial ?

À l'issue de notre exploration, qu'avons-nous appris des RCT sur le microcrédit dans le domaine du développement ? Pour revenir au titre de ce chapitre, si le microcrédit n'est pas un miracle, comme le défend le numéro spécial, que sont

les RCT sur le microcrédit : *miracle ou mirage* ? Nous allons faire le point sur nos résultats et proposer quelques pistes de réflexion.

Nous commencerons par aborder les questions de validité interne, le point fort des RCT salué par tous. Premièrement, comme l'admettent les *randomistas* eux-mêmes, il n'y a pas de preuves solides que le microcrédit est transformateur, tout comme il n'y a pas de preuves solides qu'il ne l'est pas (BANERJEE *et al.*, 2015c). Étant donné que les RCT pâtissent généralement d'un manque de puissance statistique en raison du faible taux d'adhésion et de conformité, nous ne pouvons pas nous prononcer. Deuxièmement, comme l'admettent également les *randomistas*, les effets hétérogènes sont peut-être la norme. Le microcrédit peut être transformateur pour certains et pas pour d'autres (ou pire, le microcrédit peut être négativement transformateur). Ici encore, comme les RCT pâtissent généralement d'un manque de puissance statistique en raison du faible taux d'adhésion et de conformité, nous ne pouvons pas nous prononcer. En outre, nous ne savons pas pourquoi certains peuvent tirer profit du microcrédit et d'autres pas (ou peuvent subir une pénalité « transformatrice »). Nous n'avons aucune idée des canaux par lesquels le microcrédit pourrait avoir un impact. Troisièmement, la piètre qualité des données et les erreurs de mesure peuvent conduire à reconsidérer certains des résultats pris jusque-là pour acquis. À cet égard, les nombreux problèmes que nous avons identifiés avec la RCT marocaine méritent d'être pris au sérieux. La RCT marocaine est peut-être un cas unique (le mouton noir), mais, en ce cas, ses conclusions devraient être définitivement révoquées. Cela aurait deux répercussions directes. La démonstration globale serait affaiblie. L'échantillon « assez représentatif » utilisé pour tirer des conclusions générales deviendrait « moins représentatif ». Ses bonnes propriétés mises en avant dans le numéro afin d'estimer les problèmes d'entraînement et prévoir le taux d'adhésion et ses stratégies d'échantillonnage pour traiter la question de la faible conformité et du manque de puissance statistique partiraient en fumée. Mais ce n'est peut-être pas un cas unique (bien qu'il y ait peu de chances que d'autres RCT soient aussi peu performantes), auquel cas nous sommes face à un problème structurel. La seule façon de le savoir serait de procéder à des répliques complètes, comme la nôtre. Nous préconisons vivement cette piste de recherche. Quatrièmement, nous avons montré que de nombreuses interprétations de l'impact du microcrédit, qui sous-tendent la théorie du changement, sont biaisées, tandis que certains impacts évidents (ou causes du faible taux d'adhésion) ne sont même pas pris en compte. En outre, d'autres problèmes génériques subsistent, comme les effets d'équilibre général, les politiques macroéconomiques, etc. (il s'agit dans les deux cas de problèmes de validité interne et externe).

Deuxièmement, la validité externe n'a jamais été le point fort des RCT. Notre évaluation ne change rien à cet état de fait. Les critiques habituelles, qu'il n'est pas utile de citer à nouveau ici, sont toujours d'actualité. La nouveauté du numéro spécial est qu'il examine en tandem différentes RCT sur le microcrédit. Cependant, l'accumulation de cas individuels ne résout pas le problème. Ce que l'on gagne à diversifier des contextes géographiques, mais hyper spécifiques,

se perd dans l'hétérogénéité croissante des traitements, des opérateurs, etc. Un type de produit particulier peut fonctionner dans un contexte, mais être inopérant dans un autre. Les modifications apportées aux produits et aux règles d'octroi ne fonctionnent pas de la même manière dans le « monde réel ». Enfin, les questions éthiques restent largement ignorées malgré des divergences importantes par rapport aux bonnes pratiques dans le domaine médical et même par rapport aux RCT sociales dans les pays développés.

Si l'on tient compte de tout cela, que reste-t-il ? Pour paraphraser BANERJEE et DUFLO (2011 : 167), cités par Labrousse (chapitre 8, ce volume), près de dix ans et des dizaines de RCT sur le microcrédit plus tard : « Malheureusement, [...] encore aujourd'hui, il y [a] en fait très peu de preuves, dans un sens ou dans l'autre, sur ces questions. Ce que [les *randomistas*] appellent des preuves ne sont en fait que des études de cas [...]. » Bien que l'on ne sache pas précisément ce qu'il reste à ce stade, il est en tout cas une chose qui s'est évanouie : l'énorme quantité d'argent et de ressources dépensées, dont une partie retirée à des alternatives et à d'autres usages. Cela vaut-il la peine de dépenser des millions de dollars pour ne publier qu'un seul article académique pour chaque RCT (tabl. 5) ? Ne serait-il pas plus utile d'affecter ces fonds au financement du système statistique public d'un pays en développement afin de collecter une quantité massive de données d'observation représentatives sur le long terme ? Si les promoteurs des RCT ont admis certaines des lacunes méthodologiques évoquées dans ce chapitre, leur réponse pour les résoudre est « Nous voulons plus de RCT ! » Or, comme les RCT n'ont pas tenu leurs promesses, du moins les promesses faites au monde entier par les *randomistas* au cours de ces deux dernières décennies, il serait tout aussi légitime de dire : « Nous ne voulons plus de RCT ! »

Cette proposition peut sembler abusive, mais le raz-de-marée randomiste a été si puissant (comme le montre la façon dont ils ont balayé le passé en ignorant toutes les études non expérimentales) qu'une petite poussée dans l'autre sens ne ferait pas de mal pour rééquilibrer les choses. Notre objectif n'est toutefois pas de discréditer la méthode RCT, mais de reconnaître sa véritable valeur en récusant le piédestal sur lequel elle trône actuellement. Plutôt que « Nous voulons plus de RCT ! », nous préconisons ce message : « Nous ne voulons plus de RCT autonomes ! » Si les RCT peuvent tout à fait rester appropriées et légitimes pour certaines politiques précisément circonscrites, elles doivent néanmoins être menées selon les règles. De plus, elles ne se suffisent jamais à elles-mêmes. Il est à la fois nécessaire et possible d'utiliser d'autres méthodes sans compromettre la rigueur scientifique. Comme nous l'avons vu ici, ce pluralisme devrait être une exigence, notamment pour compléter les RCT en les contextualisant, tant avant la collecte des données que pour leur analyse. Le pluralisme est aussi une exigence pour l'ensemble des questions, projets et politiques de développement qui ne se prêtent pas aux RCT, et le microcrédit, avec ses interventions relativement bien ciblées, en est un bon exemple au vu du faible taux d'adhésion et de la complexité de ses effets. Malheureusement, pour de nombreux promoteurs des RCT, et le J-PAL en particulier, « les RCT

ne sont pas seulement en tête de liste des méthodes approuvées, pour eux, il n'y a rien d'autre sur la liste », (Ravallion, chap. 1, ce volume).

Remerciements

Nous remercions les participants de l'atelier de mars 2019, qui a rassemblé la plupart des contributeurs à l'ouvrage, ainsi que Solène Morvant-Roux, Jonathan Morduch et Martin Ravallion pour leurs commentaires sur une version antérieure du chapitre.

Expérimentations aléatoires dans le champ du développement

Une perspective critique

Éditeurs scientifiques

Florent BÉDÉCARRATS, Isabelle GUÉRIN et François ROUBAUD

IRD Éditions

INSTITUT DE RECHERCHE POUR LE DÉVELOPPEMENT

Collection Synthèses

Marseille, 2022

Ce texte a fait l'objet d'une première édition en anglais :

BÉDÉCARRATS F., GUÉRIN I., ROUBAUD F. (eds), 2020, Randomized Control Trials in the Field of Development : a Critical Perspective, Oxford, Oxford University Press, 448 p.

<https://global.oup.com/academic/product/randomized-control-trials-in-the-field-of-development-9780198865360>

Photo 1^{re} et 4^e de couverture

© Ermell – Paul Klee, *Klippen am Meer*, 1931, Musée Lenbachhaus, Munich.

https://upload.wikimedia.org/wikipedia/commons/6/69/Klee_Klippen_am_Meer_1290074.jpg

Coordination éditoriale

IRD/Romain Costa

Préparation éditoriale

Marie-Laure Portal (11)

Mise en page

Desk (53)

Maquette de couverture

IRD/Michelle Saint-Léger

Maquette intérieure

IRD/Pierre Lopez



© IRD, 2022

Certains droits réservés. Il s'agit d'une publication en open access, disponible en ligne et distribuée sous les termes de l'attribution Creative commons, non commerciale, non modifiable 4.0 licence internationale (CC BY-NC-ND 4.0), dont une copie est disponible à cette adresse :

<http://creativecommons.org/licences/by-nc-nd/4.0/>

ISBN papier : 978-2-7099-2947-9

ISSN : 2431-7128

ISBN PDF : 978-2-7099-2948-6

ISBN epub : 978-2-7099-2949-3