

# La méthode d'évaluation aléatoire : apports et limites

PHILIPPE ZAMORA

L'usage de la méthode aléatoire dans le cadre de l'évaluation des politiques publiques se développe aujourd'hui plus que jamais<sup>1</sup>. La France est l'un des pays où cette méthode a connu un essor aussi vif que récent, sous l'impulsion notamment du haut-commissariat aux Solidarités actives et à la Jeunesse. Précisons notamment que cette méthode n'est pas nouvelle, elle a d'abord été introduite par les sciences médicales et dans l'évaluation des médicaments. Aujourd'hui, aucun nouveau médicament n'est mis sur le marché sans être d'abord éprouvé par un dispositif d'évaluation aléatoire. Dans le domaine des sciences sociales, elle a été utilisée depuis au moins une quarantaine d'années aux États-Unis, dans le cadre de l'évaluation des programmes sociaux (programmes de formation, programmes de tutorat, programmes d'accompagnement, notamment) et d'éducation, dans une moindre mesure. Elle a également été utilisée avec profit en économie du développement, sous l'impulsion, notamment d'Abjiit Banerjee, d'Esther Duflo et de Sendhil Mullainhatan (Banerjee et Duflo, 2009) et des membres du laboratoire J-PAL (Abdul Latif Jameel - Poverty Action Lab) créé en 2003 par ces trois économistes au MIT (Massachusetts Institute of Technology, Boston). Cette courte présentation se propose de rappeler les principes fondamentaux et les inestimables apports de cette méthode, d'appeler les limites et les précautions importantes dont elle doit s'accompagner. Les aspects éthiques sont également abordés<sup>2</sup>.

1 Je remercie Luc Behaghel pour son avis sur une première version de cet article.

2 Pour une présentation très complète mais plus technique, voir Duflo *et al.* (2007).

## Principes et apports

### *La plus-value d'une politique publique*

Généralement, lorsqu'une politique est mise en place, les décideurs et les administrations responsables sont démunis pour illustrer les résultats des programmes mis en place autrement qu'en observant le nombre de personnes passées par le programme. Ils sont en général incapables d'aller plus loin, car ils ont bien d'autres problèmes à régler (juridiques, politiques, administratifs et pratiques) : il est très rare qu'ils soient appuyés par des services statistiques ou économiques spécialisés au moment de la mise en place du programme. Les administrations opérationnelles diffusent le nombre de personnes passées dans celui-ci et dans le meilleur des cas les moyennes de  $Y$  pour les individus bénéficiaires du programme.

En réalité, lorsque l'on introduit un programme, c'est en général la question de la plus-value qui est posée. Quelle est la plus-value du programme pour ses bénéficiaires? Appliquée à une variable de résultat  $Y$ , la plus-value ( $PV$ ) désigne la différence entre la moyenne de  $Y$  pour les bénéficiaires ( $Y^?$ ) et la moyenne de  $Y$ , pour ces mêmes bénéficiaires, s'ils n'étaient pas entrés dans le programme ( $Y^0$ ).

L'inconvénient de ce concept est qu'il n'est bien sûr pas calculable directement. Si le premier terme (la moyenne de  $Y$  pour les bénéficiaires) est disponible, la deuxième partie de la différence est virtuelle, puisqu'il n'est pas concrètement possible d'observer  $Y^0$  pour les bénéficiaires effectifs du programme. Pourtant, chaque décideur ou observateur a ce concept en tête lorsqu'il cherche – même inconsciemment – à déterminer la plus-value du programme, et ce même lorsqu'il ne dispose que du premier terme de la différence. Il cherche à comparer les performances observables à celles d'une population de référence (aussi identique que possible). Comme généralement ce terme de référence n'est pas disponible, chaque observateur interprète les résultats obtenus sur  $Y$  en fonction de ses propres attentes, de sa propre connaissance du dispositif, du public spécifique. Les résultats ne sont donc pas objectivables. Cela conduit à des perceptions diverses et non réellement formalisables de la plus-value du dispositif. C'est pourquoi les observateurs ont généralement du mal à s'accorder sur l'intérêt de tel ou tel dispositif, et l'efficacité des programmes sociaux donne lieu à d'interminables débats. L'évaluation quantitative d'un programme nécessite qu'on se mette en capacité de calculer ce deuxième terme. Cette opération est loin d'être simple.

*Les évaluations ex post*

Pour répondre à ce besoin, les services statistiques ou les organismes de recherche conduisent des enquêtes *ex post*. Ces enquêtes suivent non seulement le devenir des bénéficiaires mais aussi des non-bénéficiaires. Généralement les non-bénéficiaires interrogés sont tirés au sort parmi la population qui n'est pas rentrée dans le programme. Les responsables des enquêtes cherchent à constituer des échantillons de non-bénéficiaires qui disposent de caractéristiques observables  $X$  aussi proches que possible des bénéficiaires. Dans les cas du panel des bénéficiaires des politiques d'emploi mis en place par la Direction de l'animation de la recherche, des études et des statistiques (DARES) par exemple, elles sont constituées du dernier diplôme obtenu, de l'âge, du sexe, de l'ancienneté au chômage. Le terme de référence  $Y^0$  est ici formulé : il est calculable comme la moyenne de  $Y$  des non-bénéficiaires, qui présentent un profil moyen exactement similaire au groupe des bénéficiaires. La plus-value est donc objectivée et calculable. Cette méthode constitue une première méthode de calcul explicite de la plus-value du programme.

Néanmoins, cette méthode suppose que les caractéristiques observables par enquête permettent de résumer le profil pertinent d'une personne et conditionnent quasi complètement son avenir professionnel. Cette hypothèse est remise en cause généralement par la littérature économique sur la question. En réalité, il existe de nombreuses caractéristiques – peu mesurables voire totalement inobservables par enquête statistique – qui ont un effet direct sur l'avenir professionnel d'une personne et qui conditionnent son entrée dans le programme évalué. Il est hautement improbable que des individus tirés au sort parmi les non-bénéficiaires puissent être considérés comme identiques aux bénéficiaires, du point de vue de leur devenir professionnel spontané. Si le calcul de  $Y^0$  ignore ces caractéristiques inobservées, il y a donc de forts risques de sous- ou surestimation de l'effet réel des programmes. On appelle cela le biais de sélection.

Il existe tout un courant de la recherche économétrique qui met en œuvre des méthodes complexes pour corriger de tels biais de sélection. Ces méthodes sont très techniques et reposent parfois sur des hypothèses fortes, ou bien donnent lieu à des résultats tellement imprécis qu'elles ne peuvent pas fournir des lignes directrices précises aux décideurs. En tout état de cause, il n'y a pas de consensus clair parmi les spécialistes sur les méthodes les plus performantes.

*Les évaluations aléatoires*

Pour éviter à coup sûr ce biais de sélection, une méthode de plus en plus utilisée est d'intercaler dans le processus de sélection  $D$  un tirage aléatoire. Ce tirage va déterminer deux groupes : un groupe orienté et un groupe de contrôle. Le principe de base est simple : il s'agit de faire en sorte que le groupe de contrôle soit orienté vers le programme dans une moindre proportion que le groupe orienté.

Pourquoi une évaluation aléatoire résout-elle le problème du biais de sélection ? Ce résultat repose sur la loi statistique des grands nombres. Lorsque deux échantillons  $T_1$  et  $T_2$  sont tirés au sort dans une même population-mère, les moyennes *théoriques* de n'importe quelle variable  $X$  calculée sur ces deux échantillons sont égales à la moyenne théorique de  $X$ . Autrement dit, si les effectifs des deux sous-échantillons sont suffisamment grands, ils auront spontanément le même devenir. En l'absence de toute action différenciée sur  $T_1$  et sur  $T_2$ , les différences statistiques sur n'importe quelle variable présente, future ou passée seront statistiquement nulles (c'est-à-dire à une erreur statistique près dont la distribution est contrôlée et connue). Toute la puissance de l'évaluation aléatoire réside dans ce principe : si une action est faite sur  $T_1$  et non sur  $T_2$ , alors toute différence statistique observable entre  $T_1$  et  $T_2$  (significative, c'est-à-dire suffisamment grande et qui s'éloigne significativement de la distribution spontanée de l'erreur spontanée) pourra être attribuée causalement à cette action.

Cette méthode permet de résoudre les biais de sélection et de quantifier l'effet causal pur du programme sur un certain nombre de variables d'intérêt. Les moyens techniques sont simples : il s'agit du calcul de moyennes et d'écart-type. Les hypothèses de base sont extrêmement réduites. Les résultats sont aisément lisibles pour des observateurs qui ne sont pas des spécialistes du champ statistique. L'intérêt est donc à la fois scientifique et politique. Il permet non seulement de résoudre un problème technique important mais aussi d'ouvrir un espace de communication et de coopération potentiellement extrêmement fécond entre scientifiques, praticiens et décideurs. Il permet de transporter le débat de la méthode vers le fond et ouvre les discussions sur l'interprétation plutôt que sur les hypothèses techniques.

Il faut bien prendre en compte qu'une grande part de l'organisation d'une évaluation aléatoire doit être consacrée à la construction des données. Il n'est généralement pas simple d'observer le devenir des participants à une expérimentation, car il faut souvent mettre en place un dispositif de collecte d'information *ad hoc*. Il convient, de plus, que ce dispositif atteigne des taux de réponse au moins supérieurs à 70 % pour que les résultats soient utiles.

À cette fin, il est nécessaire d'y consacrer un effort important et potentiellement coûteux.

Il s'agit là d'une tâche à laquelle les économètres académiques (pour l'instant la plupart des projets émanent de laboratoires de ce champ disciplinaire) ne sont pas forcément préparés. De plus, il faut lui consacrer des moyens importants et c'est ce poste de dépenses qui explique la majeure partie du coût d'une évaluation aléatoire. Mais ce processus d'évaluation place les chercheurs dans une position beaucoup plus avantageuse qu'en temps habituel. Le chercheur responsable de l'évaluation est en mesure de construire des données totalement adaptées aux questions qu'il cherche à résoudre. La maîtrise des données lui permet également d'explorer des terrains jusque-là peu défrichés par les enquêtes statistiques émanant de l'appareil statistique officiel et de faire se rencontrer les expériences empiriques de divers champs disciplinaires. En particulier, un grand nombre d'études aléatoires cherchent à mesurer l'effet des programmes sociaux non seulement sur des actes mais également sur des aspects subjectifs (l'estime de soi, la motivation, les attentes, projets), et intègrent des échelles mises en œuvre généralement par les psychologues cognitivistes (échelles de personnalité, etc.).

### *Une dynamique récente*

En France, les évaluations aléatoires ont connu un renouveau certain au cours de ces trois dernières années. L'exemple le plus frappant est certainement l'évaluation récente de l'accompagnement renforcé des chômeurs, lancée par l'UNEDIC et l'ANPE en 2007. Il s'agissait d'évaluer par la méthode aléatoire les conséquences de l'externalisation à des agences de placement privées, et du dispositif Cap vers l'entreprise (CVE) mis en œuvre par l'ANPE, sur le retour à l'emploi durable de chômeurs de longue durée. Cette évaluation a été mise en œuvre par une équipe de l'École d'économie de Paris et du CREST (Behaghel, Crépon, Gurgand, 2009) ; 200 000 demandeurs d'emploi ont ainsi transité par ce dispositif d'évaluation. Les résultats ont été récemment mis à disposition du public. Nul ne sait ce que fera Pôle emploi de ces résultats, mais force est de constater qu'ils ne peuvent être aujourd'hui ignorés. Par la simplicité de la méthode, ils participent de fait du débat public et sont susceptibles d'influencer les décisions qui seront prises dans l'avenir par cet opérateur.

L'effort le plus massif d'impulsion donné récemment en France à ces méthodes est à mettre au compte de Martin Hirsch, lorsqu'il était haut-commissaire aux Solidarités actives et à la Jeunesse. Près de trois appels à projets ont ainsi été publiés sur des thèmes aussi divers que la lutte contre

la discrimination, l'insertion des jeunes... Un très grand nombre de projets ont ainsi été financés, dont une part relativement importante d'évaluation à caractère aléatoire. Cette méthode tend à se diffuser à travers divers laboratoires de recherche, dont certains n'étaient pourtant pas spécialisés dans l'évaluation quantitative des politiques publiques. Cet élan récent en révèle également l'intérêt politique. Elle constitue potentiellement une méthode de gouvernance qui permet de mieux objectiver les décisions et donc de mieux les faire accepter. En tout état de cause, d'autres pays pratiquent beaucoup plus ce type de méthodes. Néanmoins, ces apports ne peuvent être réels que si un certain nombre de conditions sont réunies. Les évaluations aléatoires ne peuvent donc être improvisées, d'autant plus que leur coût peut parfois être extrêmement important.

### *Les questions éthiques*

Comme les évaluations aléatoires nécessitent une intervention humaine sur les trajectoires d'êtres humains, il est absolument nécessaire que les questions éthiques soient prises soigneusement en compte<sup>3</sup>. Il n'existe pas pour l'instant en France, à notre connaissance, de cadre juridique spécifique aux évaluations aléatoires en sciences sociales. Comme les États-Unis disposent d'une pratique de plus de quarante ans en la matière, ce sont les pratiques en vigueur dans ce pays qui guident les réflexions éthiques des concepteurs d'expérimentation en France. Le rapport Belmont rédigé en 1979 a fixé le cadre de ces pratiques. Celles-ci reposent sur trois principes éthiques :

1. *Respect des personnes* : traiter les individus comme des personnes autonomes et protéger particulièrement les personnes dont l'autonomie est limitée (enfants, prisonniers, personnes handicapées par exemple).
2. *Bienfaisance* : il s'agit d'apprécier les risques que courent les personnes en expérimentation par rapport aux bénéfices collectifs attendus.
3. *Justice* : il est important que les avantages de la recherche bénéficient à tous et que les sujets expérimentés ne soient pas systématiquement des sujets plus faibles et plus simples à mobiliser.

Il s'agit là de principes généraux qui doivent être interprétés au cas par cas et qui donnent lieu à une jurisprudence volumineuse. Chaque projet envisagé doit, aux États-Unis, être examiné par un comité d'éthique spécialisé dans la recherche en sciences sociales, présent dans chaque université. Ces comités examinent tout projet qui fait intervenir une action sur des personnes humaines – ce qui va bien au-delà des évaluations aléatoires. Par exemple,

---

3 Voir, dans ce numéro, la contribution de Cécile Campy, pour une illustration pratique de ces enjeux éthiques, dans un cas d'évaluation aléatoire.

tout projet d'enquête – si anodin qu'il soit – est soumis en premier lieu à un comité d'éthique. Ces comités sont composés de personnalités qualifiées qui doivent être issues de diverses disciplines (en plus des sciences humaines et sociales). Ainsi, la plupart des comités d'éthique sont composés de philosophes, de juristes et de médecins. En France, de tels comités n'existent pas : seule l'École d'économie de Paris (qui abrite le J-Pal, spécialisé dans les évaluations aléatoires) a créé récemment le sien. On pourra se reporter au site du comité d'éthique du MIT pour des informations détaillées<sup>4</sup>.

Le consentement éclairé est l'un des sujets de discussion les plus importants. En sciences médicales, il est absolument obligatoire. Aucun patient ne peut être soumis au test expérimental d'un médicament ou d'une pratique médicale à son insu. Il doit choisir d'y participer en pleine connaissance de cause et en toute information des risques encourus. Toutefois, l'introduction d'un tel cadre est plus problématique en sciences sociales. Deux problèmes se posent en effet.

Premièrement, à l'inverse des sciences médicales, il est généralement impossible de conduire des évaluations en aveugle, c'est-à-dire d'avoir conscience de faire partie d'une expérimentation tout en ne sachant pas si l'on fait partie du groupe de contrôle ou du groupe traité. En sciences sociales, les dispositifs en aveugle sont plus difficiles à monter : maquiller un dispositif de droit commun (différent du dispositif à évaluer), et faire accroire qu'il s'agit d'un dispositif expérimental, contreviendrait aux règles éthiques les plus élémentaires.

Deuxièmement, le sentiment de faire partie d'une évaluation expérimentale risque de modifier à la fois le comportement des groupes traités et des groupes de contrôle (dans des directions potentiellement différentes). En conséquence, cela est susceptible d'altérer la validité externe de l'évaluation : les résultats obtenus risquent en effet d'être absolument peu conformes à la situation qui prévaudrait en cas d'implantation permanente du dispositif évalué.

C'est pourquoi le rapport Belmont prévoit la possibilité de dérogations à l'instauration des consentements éclairés. Mais cela nécessite un examen d'autant plus approfondi des risques encourus par les groupes traités et les groupes de contrôle et une surveillance accrue du devenir des individus des deux groupes par les équipes des expérimentateurs. Tout doit être fait pour réagir rapidement si l'un des traitements s'avère néfaste pour un individu faisant partie d'un des deux groupes expérimentaux.

4 Committee on the Use of Humans as Experimental Subjects (COUHES), [URL : <http://web.mit.edu/committees/couhes/>], consulté le 15 avril 2011.

Il convient de préciser cependant qu'il existe une particularité propre aux évaluations dans le domaine des sciences sociales. Généralement, la plupart des programmes évalués sont expérimentaux : ils bénéficient de financements spécifiques, juste suffisants pour accueillir une proportion infime de la population qui pourrait en bénéficier (en cas de généralisation). Le nombre de personnes éligibles au dispositif après diagnostic excède donc largement le nombre de personnes que le programme est en capacité de recevoir. Face à ce rationnement, parmi tous les individus susceptibles d'être choisis, il est difficile d'être sûr que tel ou tel individu mérite d'intégrer le dispositif davantage qu'un autre. En l'absence d'informations spécifiques, il est finalement moins injuste de tirer au sort les individus qui vont être retenus que de les désigner par l'application d'une expertise humaine (ou a fortiori politique). Compte tenu de l'apport collectif possible, qui doit être jugé collectivement (par un comité d'éthique compétent et composé avec diversité), la sélection aléatoire – si elle est conduite de façon contrôlée et transitoire – peut donc être un choix juste et souhaitable.

*Limites et précautions. Les rapports entre évaluateurs et acteurs de terrain*

Une évaluation à caractère aléatoire nécessite une intervention *ex ante*, au cœur même du processus de sélection. Elle vient s'immiscer au sein-même du travail des praticiens, mais elle doit s'efforcer de le modifier le moins possible, sinon pas du tout. L'instrument de mesure ne doit, bien sûr, pas influencer ce qu'il cherche à cerner. Il convient aussi qu'il y ait une bonne adhésion des praticiens à l'objectif de l'évaluation. Cette adhésion doit s'appliquer à trois aspects : 1) l'objectif politique de l'évaluation, 2) les questions éthiques qu'elle soulève, et enfin 3) les contraintes pratiques qu'elle provoque.

1. À l'inverse des méthodes d'évaluation statistique *ex post*, les évaluations aléatoires reposent sur très peu d'hypothèses. Elles ont pour elles une très grande simplicité conceptuelle et technique : les résultats sont immédiatement compréhensibles par les décideurs. De ce fait, les résultats sont clairs, neutres et difficiles à contester ou à relativiser. Leurs implications politiques potentielles en sont donc amplifiées. Ainsi, aux États-Unis, les évaluations aléatoires des grands programmes fédéraux de formation et d'accompagnement ont conduit à des réaménagements très importants des budgets alloués (LaLonde, 1986). La présence d'enjeux financiers et humains importants peut donc générer certaines résistances à l'égard de ces évaluations et celles-ci peuvent se révéler d'autant plus contre-productives que la conduite de telles opérations nécessite une coopération étroite et permanente entre évaluateurs et acteurs de terrain.

2. Une des sources de conflits potentiels entre évaluateurs et acteurs de terrain réside dans une appréhension différente des questions éthiques. Même lorsque le dispositif est expérimental et/ou qu'il y a plus de candidats éligibles que de places, les acteurs de terrain sont souvent confrontés à la difficulté de conduire et justifier le tirage au sort auprès des candidats potentiels. Même si celui-ci est théoriquement justifié et a été soumis aux instances compétentes, sa place et sa réception sur le terrain sont une autre affaire. Certaines personnes restent réfractaires à ce principe méthodologique pour des raisons philosophiques et restent peu perméables aux arguments qui le fondent. En général, ils n'en partagent pas les présupposés. Par nature, l'évaluation est agnostique : elle questionne entièrement, à nouveaux frais, l'efficacité du programme. Ses principes éthiques procèdent de cette neutralité fondamentale : en l'absence d'évidence sur le programme, il n'est pas injuste d'en tirer au sort les participants effectifs et d'orienter les membres du groupe de contrôle vers d'autres dispositifs existants. Or les acteurs de terrain peuvent par leurs propres moyens d'expertise, leur expérience, ou leurs croyances être convaincus de l'efficacité de leur programme. Généralement, lorsqu'un dispositif est complètement expérimental, qu'il est créé et pensé d'emblée comme un programme à évaluer, ce décalage est faible et ces réticences éthiques peuvent être facilement résolues si on prend le temps de discuter posément avec chacun des acteurs concernés. Mais, lorsqu'un dispositif est installé, préexiste depuis plusieurs années, il devient très difficile de justifier le tirage au sort, alors même que les conditions théoriquement nécessaires sont réunies.

3. Introduire un tirage au sort dans le processus de sélection des bénéficiaires n'est pas forcément simple. Il est difficile de dégager des idées générales tant cela dépend des conditions concrètes. En tout état de cause, il est par définition très difficile de modifier des conditions de fonctionnement préétablies, et ce d'autant plus que l'évaluation se déroule au sein d'institutions de grande taille ou dont les lignes hiérarchiques sont mal définies ou complexes. Il s'agit bien sûr d'une question d'inertie des organisations, mais cette difficulté peut tenir également à certaines réticences compréhensibles : les acteurs peuvent se sentir dépossédés de leur expertise ou de leur rôle lorsqu'à un moment ou à un autre un processus de tirage au sort vient perturber les procédures habituelles. Il est nécessaire d'intégrer ces aspects le plus en amont possible et, par exemple, de prévoir des incitations pour que les organisations puissent s'adapter plus facilement aux exigences de la méthode aléatoire.

### *La validité externe*

Pour que les résultats des évaluations aléatoires soient réellement utiles aux décideurs, il convient que les conditions de l'expérimentation soient

proches de celles qui prévaudraient en cas d'instauration permanente du dispositif. En particulier, les individus ne doivent pas modifier leur comportement du fait même qu'ils participent à l'expérimentation et qu'ils ont conscience d'être observés. Il s'agit là d'un point essentiel et c'est pourquoi l'introduction d'un consentement éclairé peut être de nature à altérer la validité externe des expérimentations sociales, à l'inverse des évaluations dans le cadre médical. Il convient également que le public entrant dans l'expérimentation ne soit pas trop différent de celui qui entrera ensuite en cas d'adoption du programme. Par exemple, si l'on souhaite expérimenter un nouveau mode d'accompagnement et le généraliser, le réserver aux seuls volontaires pendant l'évaluation risque de biaiser considérablement les résultats souhaités.

Les décideurs demandent souvent quelle est la plus-value ou l'efficience (rapport coût-bénéfice) du dispositif évalué en cas de généralisation. Il s'agit là d'une question difficile, car l'efficacité constatée lorsqu'une partie infime de la population passe par le programme peut être complètement bouleversée si celui-ci est généralisé à l'ensemble de la population éligible. Il ne s'agit pas ici simplement d'éventuels effets de composition de la population, mais du fait que les mécanismes économiques fassent souvent intervenir des « effets de bouclage » macroéconomiques, dont l'évaluation expérimentale ne peut tenir compte spontanément ou qu'elle ne peut mesurer facilement. À cet égard, on peut évoquer un exemple classique concernant les effets potentiels à court terme de l'accompagnement renforcé. Si l'on suppose que les emplois constituent un stock peu flexible à court terme, il est possible d'observer un effet à la fois positif de l'accompagnement sur le taux de chômage des personnes accompagnées, lorsqu'elles sont en faible nombre, et rigoureusement nul si l'accompagnement renforcé est généralisé à toutes les personnes éloignées de l'emploi. En effet, dans le cadre des hypothèses restrictives de ce modèle, l'accompagnement a dans ce cas simplement un rôle de modification de la file d'attente. Au total, le taux de chômage global n'est pas sensible à cette politique.

Il existe très peu de travaux qui cherchent à mesurer directement les effets de bouclage (ou effets externes) appliqués aux programmes sociaux, bien que la littérature théorique s'y soit intéressée depuis de longues années (Cahuc et Le Barbanchon, 2010). À ce jour, les quelques articles empiriques existants sur des programmes sociaux en Europe ou aux États-Unis n'ont pas mesuré d'effets externes significatifs. Un travail récent a été lancé en France pour se mettre en capacité d'évaluer ces effets de bouclage sur l'accompagnement renforcé des jeunes de l'enseignement supérieur (Crépon *et al.*, 2010).

*La précision*

Comme dans tout exercice statistique, les résultats obtenus sont entachés d'incertitude. Celle-ci dépend du nombre d'individus qui entrent dans l'expérimentation, mais également de la différence entre le taux d'entrée dans le programme évalué des personnes du groupe traité et du groupe de contrôle respectivement. Plus cette différence est grande, moins il faut de personnes expérimentées pour obtenir une précision donnée sur l'effet évalué.

Il s'agit là d'un point extrêmement important : les évaluations *ex post* se réalisent souvent en analyse secondaire et les chercheurs sont souvent dépendants des conditions d'enquête fixées par d'autres qu'eux-mêmes ; à l'inverse, dans les évaluations aléatoires *ex ante*, les personnes responsables ont souvent la possibilité de fixer au préalable le *design* de l'expérimentation et de calculer un « effet minimum détectable ». Cet effet est un concept statistique : il est calculé *ex ante* et incorpore les différentes distributions d'erreur statistique qui se réaliseront potentiellement. Il est possible que, *ex post*, l'effet réellement mesurable se révèle plus faible que cet effet calculé *ex ante*.

Le calcul *ex ante* des effets minimaux détectables (avant le lancement effectif du protocole aléatoire) n'est pas forcément un exercice très simple, il est même parfois nécessaire de conduire des premiers tests-pilotes pour déterminer correctement les différents paramètres. En tout état de cause, c'est un exercice nécessaire, car il convient de ne pas engager des coûts parfois importants pour être en définitive incapable de porter un diagnostic suffisamment précis sur le programme évalué et ainsi décevoir les attentes suscitées. En conséquence, il convient toujours de formuler le plus clairement possible un effet escompté et de définir le *design* de l'expérimentation en fonction de celui-ci. Lorsque les effets escomptés sont trop faibles pour être mesurés dans les conditions matériellement possibles, il est alors préférable de renoncer à l'expérimentation.

## Bibliographie

- BANERJEE Abhijit et DUFLO Esther, 2009, « The experimental approach to development », *Annual Review of Economics*, vol. I, n° 1, p. 151-178.
- BEHAGHEL LUC, CRÉPON BRUNO, et GURGAND MARC, 2009, « Évaluation d'impact de l'accompagnement des demandeurs d'emploi par les opérateurs privés de placement et le programme Cap vers l'entreprise. Rapport final » [en ligne], Rapport pour l'Agence nationale pour l'emploi, Paris, [URL : [http://www.crest.fr/ckfinder/userfiles/files/Pageperso/crepon/Rapport\\_final\\_Crest\\_EEP.pdf](http://www.crest.fr/ckfinder/userfiles/files/Pageperso/crepon/Rapport_final_Crest_EEP.pdf)], consulté le 11 avril 2011.

- CAHUC Pierre et LE BARBANCHON Thomas, 2010, « Labor market policy evaluation in equilibrium : some lessons of the job search and matching model », *Labour Economics*, vol. 17, n° 1, p. 196-205.
- CRÉPON Bruno, DUFLO Esther, GURGAND Marc, RATHELOT Roland et ZAMORA Philippe, 2010, « Le marché d'accompagnement des jeunes diplômés demandeurs d'emploi par des opérateurs privés de placement : premiers résultats » [en ligne], [URL : <http://www.travail-emploi-sante.gouv.fr/IMG/pdf/100103-noteintermediaire.pdf>], consulté le 20 mars 2011.
- DUFLO Esther, GLENNESTER Rachel et KREMER Michael, 2007, « Using randomization in development economics research : a toolkit », *Handbook of Development Economics*, vol. 4, p. 3895-3962.
- LALONDE Robert J., 1986, « Evaluating the econometric evaluation of training programs using experimental data », *American Economic Review*, vol. 76, n° 4, p. 604-620.
- LHORTY Yannick et PETIT Pascale, 2010, « Évaluations aléatoires et expérimentations sociales », *Document de recherche EPEE*, n° 2010-2, Université d'Évry.