



HAL
open science

L'évaluation socioéconomique de l'investissement social

Denis Fougère, Arthur Heim

► **To cite this version:**

Denis Fougère, Arthur Heim. L'évaluation socioéconomique de l'investissement social : Comment mettre en oeuvre des analyses coûts-bénéfices pour les politiques d'emploi, de santé et d'éducation. [Rapport de recherche] France Stratégie - Commissariat général à la stratégie et la prospective. 2019, pp.184. hal-03456048

HAL Id: hal-03456048

<https://sciencespo.hal.science/hal-03456048>

Submitted on 29 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

L'évaluation socioéconomique de l'investissement social

Comment mettre en œuvre des analyses
coûts-bénéfices pour les politiques
d'emploi, de santé et d'éducation

Denis Fougère
Arthur Heim



FRANCE STRATÉGIE
ÉVALUER. ANTICIPER. DÉBATTRE. PROPOSER.



FRANCE STRATÉGIE
ÉVALUER. ANTICIPER. DÉBATTRE. PROPOSER.

L'évaluation socioéconomique de l'investissement social

Comment mettre en œuvre des analyses coûts-bénéfices
pour les politiques d'emploi, de santé et d'éducation

Document de travail

Denis Fougère et Arthur Heim

Novembre 2019

Sommaire

Préambule	5
------------------------	----------

Chapitre 1 – Les raisons et principes de l'application du calcul socioéconomique aux politiques d'investissement social

1. Du développement des mesures d'impact et du calcul socioéconomique	11
2. Qu'est-ce qui distingue le calcul socioéconomique appliqué à l'investissement social des autres formes d'évaluation ?	15
3. Dans quelle phase du processus évaluatif doit-on tenir compte des bénéfices et des coûts anticipés d'une politique d'investissement social ?	19

Chapitre 2 – Définition et déroulé de l'évaluation d'une politique d'investissement social

1. Cadrage général	23
2. Adapter les méthodes d'évaluation aux questions posées	26
3. Les étapes de la mise en œuvre du calcul socioéconomique	27
4. Quelles études sélectionner pour une revue de littérature ?	37

Chapitre 3 – Le cadre général de l'application du calcul socioéconomique aux politiques d'investissement social

1. La valeur actualisée nette	47
2. Les principaux paramètres	48
3. Comment convertir des effets intermédiaires et prendre en compte les effets inobservés (connus)?	59
4. Quel horizon temporel et quelles données privilégier ?	63
5. Le ratio coûts-efficacité	65
6. Les variables de contexte et le scénario de référence	68
7. Les effets induits	69
8. Situation contrefactuelle et expérimentation	74
9. Recension et quantification des coûts de l'intervention	82

Chapitre 4 – Prolongements et spécificités du calcul socioéconomique appliqué à l'investissement social

1. Évaluations <i>ex-post</i> ou <i>ex-ante</i> , parcimonie ou complexité	85
2. Identifier la valeur sociale d'un résultat à l'aide d'un prix implicite	91

3. Équité et redistribution dans les analyses coûts-bénéfices	100
4. Les effets d'équilibre général.....	106
Résumé des conclusions.....	121
Recommandations	123
Bibliographie.....	127
Annexes A – Mise en pratique d'analyses coûts-bénéfices d'investissements sociaux	139
Annexe A-I – Le programme STAR de réduction de la taille des classes	140
Annexe A-II – Les analyses coûts-bénéfices du Washington State Institute for Public Policy.....	145
Annexe A-III – Comment choisir la prévention en santé la plus efficiente ? L'exemple de la modélisation sur la population australienne	151
Annexe A-IV – Prise en compte des enjeux d'équité dans l'analyse coûts-bénéfices : une application sur données françaises	155
Annexe A-V – Le programme <i>Perry Preschool</i>	160
Annexes B – Quelques institutions produisant des revues systématiques de littérature	167
Annexe B-I – Campbell Collaboration	167
Annexe B-II – Washington State Institute for Public Policy	168
Annexe B-III Cochrane	168
Annexe B-IV Education Endowment Foundation	168
Annexe B-V What Works Clearing House	169
Annexe B-VI National institute for health and care excellence (NICE).....	169
Annexes C – Quelques références de méta-analyses	171
Annexe C-I – Quelques méta-analyses de l'impact de politiques d'emploi.....	171
Annexe C-II – Quelques méta-analyses de politiques éducatives	172
Annexe C-II – Quelques revues systématiques de politiques de prévention en santé.....	174
Annexe D – Liste des membres du groupe de travail	177
Annexe E – Déroulé des séances du groupe de travail.....	179

Préambule

Le calcul socioéconomique est l'une des missions de France Stratégie qui fait suite à une longue tradition de travaux amorcée par le Commissariat général du Plan (jusqu'en 2006) puis par le Centre d'analyse stratégique (2006-2013)¹. Les premiers travaux sur ce thème, synthétisés dans la publication des rapports Boiteux I² en 1994 et Boiteux³ II en 2001, puis du rapport Lebègue⁴ en 2005, concernaient les transports et le choix du taux d'actualisation. Le Centre d'analyse stratégique (CAS) a élargi les champs d'application aux enjeux du développement durable avec un premier rapport Quinet⁵ (2008) sur la valeur tutélaire du carbone, puis avec le rapport Chevassus-au-Louis⁶ (2009) sur la biodiversité. La dernière publication du CAS ayant contribué à cette réflexion est le rapport Gollier⁷ (2011) sur le calcul du risque dans les investissements publics.

La publication du rapport Quinet⁸ (2013) par France Stratégie marque un tournant dans l'usage du calcul socioéconomique en apportant un cadrage méthodologique pour répondre aux nouvelles exigences du législateur. En effet, la loi du 31 décembre 2012 prévoit que tout projet d'investissement financé par l'État ou ses établissements publics doit faire l'objet d'une évaluation socioéconomique préalable à son approbation.

¹ Ce document de travail a fait l'objet d'une note de synthèse, également disponible sur le site de France Stratégie : voir Fougère D. et Heim A. (2019), « [L'investissement social à l'épreuve de l'évaluation socioéconomique](#) », *La Note de synthèse*, n° 82, France Stratégie, novembre.

² Boiteux M. (1994), *Transports : pour un meilleur choix des investissements. Rapport pour le Commissariat général du Plan*, La Documentation française, Paris, 132 p.

³ Boiteux M. (2001), *Transports : choix des investissements et coût des nuisances, Rapport pour le Commissariat général du Plan*, La Documentation française, Paris, 328 p.

⁴ Lebègue D. (2005), *Révision du taux d'actualisation des investissements publics, Rapport pour le Commissariat général du Plan*, La Documentation française, Paris, 112 p.

⁵ Quinet A. (2008), *La valeur tutélaire du carbone*, Rapport pour le Centre d'analyse stratégique, La Documentation française, 424 p.

⁶ Chevassus-au-Louis B. (2009), *Approche économique de la biodiversité et des services liés aux écosystèmes*, Rapport pour le Centre d'analyse stratégique, La Documentation Française, 376 p.

⁷ Gollier C. (2011), *Le calcul du risque dans les investissements publics*, Rapport pour le Centre d'analyse stratégique, La Documentation française, 240 p.

⁸ Quinet E. (2013), *L'évaluation socioéconomique des investissements publics*, Tome 1, Rapport pour le Commissariat général à la stratégie et à la prospective, La Documentation française ; E. Quinet (2013), *L'évaluation socioéconomique des investissements publics*, Tome 2, Rapport pour le Commissariat général à la stratégie et à la prospective, La Documentation française.

Depuis 2016, France Stratégie et le Commissariat général à l'investissement (devenu depuis Secrétariat général pour l'investissement public, SGPI) ont mis en place un Comité d'experts des méthodes d'évaluation socioéconomique des projets d'investissements publics ayant pour but de créer une culture commune de l'évaluation des projets d'investissement en veillant à associer les différentes administrations concernées, mais aussi en intégrant dans les fonctions d'expert ou de rapporteur des économistes formés aux techniques d'évaluation les plus récentes. Le comité a quatre missions :

- préciser les règles méthodologiques de l'évaluation socioéconomique ;
- définir les études et recherches nécessaires à entreprendre à cette fin ;
- renforcer l'usage du calcul socioéconomique (au sein de l'État et de ses établissements mais aussi au sein des collectivités) en l'adaptant pour répondre aux difficultés rencontrées dans son application ;
- contribuer à vulgariser et promouvoir la pratique du calcul socioéconomique.

La principale contribution de ce groupe a été de produire le guide⁹ de l'évaluation socioéconomique des investissements publics et plusieurs travaux sectoriels, notamment sur le déploiement du haut débit ou sur l'évaluation des investissements dans les établissements d'enseignement supérieur. Ce groupe a également produit un tableur de valeurs tutélaires, qui couvre les champs de la santé, du climat, de l'environnement, du temps et du bruit¹⁰.

L'investissement social : une conception renouvelée de l'action publique

Depuis le début des années 2000, l'investissement social est devenu un concept central dans le choix et la mise en œuvre des politiques publiques. Cette notion, mise en avant par la Commission européenne, a sans nul doute permis de renouveler le modèle de l'action sociale. L'investissement social consiste à « investir dans les personnes en adoptant des mesures pour renforcer leurs compétences et leurs capacités, leur autonomie et leur permettre de participer pleinement au monde du travail et à la société¹¹ ». Dans cette perspective, il s'agit d'orienter l'action publique de façon à accroître la dotation des individus en capacités de production (*i.e.* en capital humain), d'adaptation et de résilience face aux risques sociaux, par opposition aux modèles de protection sociale existants plus souvent fondés sur l'assurance et la gestion des risques¹².

Cette conception renouvelée de la protection sociale — *lato sensu* incluant les politiques d'emploi, les politiques relatives à la petite enfance, les politiques éducatives, les politiques de santé, de prévention de la délinquance, de logement, ou de lutte contre la pauvreté — repose sur l'idée que les sommes dépensées aujourd'hui peuvent engendrer des bénéfices monétaires et sociaux, parfois à court terme mais aussi (et surtout) dans un futur plus

⁹ France Stratégie et DG Trésor (2017), [Guide de l'évaluation socioéconomique des investissements publics](#).

¹⁰ Ces tableurs sont accessibles [sur le site de France Stratégie](#).

¹¹ <http://ec.europa.eu/social/main.jsp?catId=1044&etlangId=fr>

¹² Jenson J. (2010), "Diffusing ideas for after neoliberalism: The social investment perspective in Europe and Latin America", *Global Social Policy*, Vol. 10(1), p. 59-84 ; Kvist J. (2015), "A framework for social investment strategies: Integrating generational, life course and gender perspectives in the EU social investment strategy", *Comparative European Politics*, Vol. 13(1), p. 131-149.

lointain, ce qui est la définition même de l'investissement. Le rendement attendu de ces politiques inclut en particulier les dépenses évitées de protection sociale¹³. Elles visent à encourager les comportements générant des externalités positives, à décourager ceux produisant des externalités négatives, mais aussi à agir sur les inégalités de revenu, de richesse ou de consommation.

La crédibilité de ce concept et la possibilité de le traduire opérationnellement dépendent fortement de l'existence de travaux académiques mesurant de façon rigoureuse les rendements de l'investissement social. Des éléments de formalisation ont été produits par quelques auteurs¹⁴, qui ont notamment utilisé des modèles dynamiques, mais cette littérature reste marginale et souvent critiquée. Les hypothèses théoriques fondant ces modèles, qu'il s'agisse de celles relatives à la croissance ou à l'emploi, sont souvent discutables, les problèmes de causalité inverse¹⁵ y sont rarement évoqués, et la distinction entre dépenses sociales d'investissement et de consommation qui y est faite est rarement convaincante.

L'investissement social : un nouveau chantier pour France Stratégie

Parallèlement aux travaux sur le calcul socioéconomique, France Stratégie a organisé durant l'année 2016 un cycle de séminaires sur l'investissement social en partenariat avec les Apprentis d'Auteuil, la Caisse nationale des Allocations familiales (Cnaf), la Direction générale de la cohésion sociale (DGCS) et le Laboratoire interdisciplinaire d'évaluation des politiques publiques de Sciences Po Paris.

Deux principaux objectifs ont guidé ces séminaires. Le premier était de clarifier le concept d'investissement social pour mieux saisir son contenu et son utilité pour l'action. Le second, de préciser les défis opérationnels de l'investissement social pour ce qui concerne la France.

Ces dimensions ont été explorées au cours de sept séances :

- une séance de lancement a dressé une comparaison entre pays et traité les enjeux généraux de définition et d'évaluation ;
- cinq séances thématiques ont traité des politiques publiques en matière de : conditions d'accueil du jeune enfant ; articulation entre vie familiale et vie professionnelle et égalité femmes/hommes ; politiques de la jeunesse ; nouvelles formes de lutte contre la pauvreté ; formation tout au long de la vie et accompagnement global vers et dans l'emploi ;
- une séance conclusive a permis à un large panel d'acteurs du champ social de se positionner par rapport à la stratégie d'investissement social.

¹³ Palier B. (2014), *La stratégie d'investissement social*. Les études du Conseil économique, social et environnemental, La Documentation française, Paris, 68 pages.

¹⁴ Voir, par exemple, pour une présentation critique de ces travaux : Heim A. (2017), « Peut-on estimer le rendement de l'investissement social ? », Document de travail, France Stratégie, Paris.

¹⁵ Le marché du travail et la croissance sont-elles plus dynamiques en raison des investissements sociaux ou bien, au contraire, les investissements sociaux sont-ils permis par un marché du travail dynamique et une croissance élevée ?

Un ouvrage¹⁶ publié en 2017 fait la synthèse des communications, débats et conclusions de ce cycle de séminaires.

En complément, France Stratégie a publié un document de travail¹⁷ et une note de synthèse¹⁸ détaillant les conditions de mise en œuvre d'évaluations du « rendement de l'investissement social ».

À la suite de ces deux publications, France Stratégie a souhaité compléter sa réflexion sur les méthodes d'évaluation socioéconomique des projets d'investissement public en constituant un groupe de travail sur l'usage du calcul socioéconomique pour l'investissement social.

Un groupe de travail sur le rendement de l'investissement social : composition et fonctionnement

À l'automne 2017, France Stratégie a sollicité Denis Fougère, directeur de recherche au CNRS (Sciences Po Paris), pour composer et présider un groupe de travail auquel trois objectifs étaient assignés :

1. clarifier, sur la base d'une recension de la littérature académique et d'études de cas pratiques, les intérêts et distinctions des calculs coûts-efficacité et coûts-bénéfices appliqués aux politiques sociales et éducatives ;
2. établir un diagnostic sur les éléments manquant aujourd'hui pour parvenir à mettre en œuvre ce type d'analyse de façon plus systématique, à la fois dans des évaluations *ex ante* et *ex post* (de quelles bases de données sur les coûts et les valeurs sociales faut-il disposer ? quelles synthèses d'évaluation sociale faut-il produire ?) ;
3. établir des recommandations pour encourager et guider la mise en œuvre d'analyses socioéconomiques en tenant compte des spécificités des interventions dans le champ social.

Le groupe de travail mis en place a réuni à France Stratégie une vingtaine d'experts issus du monde universitaire (Warwick University, Paris School of Economics, Crest, Sciences Po, Lille 2, Paris Dauphine, Toulouse School of Economics, etc.) et d'administrations et organismes publics (Drees, Santé publique France, Assurance maladie, INSEE, DG Trésor, Depp, Pôle emploi, CGI, France Stratégie).

Entre février et juillet 2018, le groupe s'est réuni six fois. Après une première séance inaugurale, les cinq réunions suivantes ont porté sur :

- les méthodes d'analyse coûts-bénéfices¹⁹, coûts-efficacité²⁰ et les techniques de microsimulation appliquées à l'investissement social ;

¹⁶ Avenel C., Boisson-Cohen M., Dauphin S., Duvoux N., Fourel C., Jullien M. et Palier B. (2017), *L'investissement social : quelle stratégie pour la France ?* La documentation française, Paris.

¹⁷ Heim A (2017), « Comment estimer le rendement de l'investissement social ? », Document de travail 2017-02, France Stratégie.

¹⁸ Heim A (2017), « Peut-on estimer le rendement de l'investissement social ? » *La note d'analyse*, n° 52, France Stratégie.

- les rendements des politiques éducatives ;
- les rendements des politiques d'emploi ;
- l'évaluation économique en santé ;
- les méthodes d'économie expérimentales et comportementales.

D'autres thèmes ont été abordés durant ces séances. Parmi ceux-ci on citera les liens entre délinquance et chômage, la lutte contre les inégalités femmes/hommes sur le marché du travail, les effets des politiques d'investissement social sur le bien-être individuel, l'équité dans les calculs socioéconomiques.

Ce groupe de travail a d'abord été pensé comme une déclinaison sectorielle des méthodes d'évaluation socioéconomiques et de la démarche proposée dans le guide de l'évaluation socioéconomique des investissements publics²¹. Cependant, à mesure que le groupe de travail progressait dans sa réflexion, l'évaluation socioéconomique – telle qu'elle a pu être définie et cadrée dans le guide précédemment cité – a dû être adaptée pour tenir compte des particularités des politiques sociales, d'éducation, d'emploi, etc., et définir ce qu'il convient alors de mettre en place.

Si, de plusieurs points de vue, la démarche générale et les étapes peuvent être comparables, l'investissement social, plus que les grands projets d'investissement public, concerne principalement des personnes et se dote d'objectifs ayant directement trait aux enjeux d'équité et de redistribution. Ces deux caractéristiques impliquent d'utiliser davantage des données individuelles et des méthodes qui permettent de caractériser les gagnants, les perdants et les valeurs en jeu. De plus, les effets d'investissements sociaux sont souvent diffus, hétérogènes en fonction des caractéristiques des personnes, et ces effets peuvent bénéficier à certains groupes mais pas à d'autres. En outre, leurs évaluations sont dépendantes de protocoles de recherche permettant d'identifier, dans les faits ou dans les données, la situation de référence, c'est-à-dire celle qui prévaut sans l'investissement social. Au contraire, les calculs socioéconomiques de grands investissements publics reposent largement sur une définition théorique ou hypothétique de la situation de référence. Pour ces raisons, dans le cas de l'investissement social, les méthodes habituelles d'évaluation socioéconomique, notamment celles reposant *ex ante* sur des modèles structurels très contraignants en termes d'hypothèses de comportement et de formes fonctionnelles, ne peuvent être appliquées telles quelles. Il est important de faire remarquer que le recours rigide aux modèles structurels est de moins en moins fréquent dans le cas des politiques d'éducation et d'emploi. Une voie de recherche fructueuse a été toutefois ouverte par plusieurs économistes qui utilisent les approches relativement athéoriques fournies par les

¹⁹ L'analyse coûts-bénéfices implique de comparer le total des coûts attendus d'une ou plusieurs politiques d'investissement social au total des bénéfices escomptés de cette ou de ces politiques, afin de déterminer laquelle est la meilleure ou la plus rentable. Le principe de l'analyse coûts-bénéfices est détaillé en plusieurs endroits de ce document.

²⁰ L'analyse coût-efficacité a pour but d'identifier la manière économiquement la plus efficace de réaliser un objectif préétabli. Cette méthode se focalise sur le résultat majeur attendu d'une politique. Elle est utilisée pour identifier l'efficacité d'un programme rapporté à son coût. Ce document précise ultérieurement cette définition et en donne des exemples.

²¹ France Stratégie, *op. cit.*

approches expérimentales et quasi expérimentales pour calibrer et simuler ensuite des variantes de modèles structurels²².

Ce que contient ce document de travail

Ce document de travail s'appuie sur les communications présentées lors des séances du groupe de travail et sur une analyse critique des études internationales rapportant à la fois les résultats d'évaluations d'impact et de mesures d'efficience calculées à l'aide d'analyses coûts-bénéfices ou coûts-efficacité.

Ce faisant, il accorde une large place à la description des étapes nécessaires à la mise en œuvre d'évaluations d'impact et des conditions permettant d'obtenir des mesures du rendement de l'investissement social. En d'autres termes, il s'agit ici de déduire la possibilité d'articuler les résultats d'évaluations rigoureuses avec des applications pertinentes du calcul socioéconomique.

Afin de clarifier les méthodes, les enjeux et les difficultés du calcul socioéconomique appliqué à l'investissement social, ce document limite son périmètre d'analyse à trois grandes familles de politiques publiques, à savoir les politiques d'éducation, d'emploi et de santé. Il n'a pas pour ambition de couvrir chacune des interventions, réformes ou dispositifs qui composent ces trois grandes familles. Il ne prétend pas être exhaustif : il vise principalement à dégager les enjeux de méthodes, à détailler les besoins en données, à indiquer les sources d'information et les techniques qui permettent de proposer un équivalent monétaire à des effets non marchands tels que la vie en bonne santé, la moindre propension à la délinquance ou à la maltraitance, la moindre dépendance aux substances psychoactives illicites, la stabilité affective et émotionnelle, etc. En un mot, on se doit de faire la liste de l'ensemble des composantes qui permettent d'évaluer des politiques sociales de la façon la plus juste et la plus complète.

À titre d'illustration, ce document présente en détail un certain nombre de travaux dont certains jouent un rôle de « cas d'école » afin d'en tirer des recommandations sur la façon de mesurer les effets directs ou induits, et dont la valorisation monétaire peut être parfois délicate.

Enfin, il insiste sur trois dimensions particulièrement importantes pour le calcul du rendement des investissements sociaux :

- les différentes méthodes de valorisation monétaire de ses effets ;
- la prise en compte de l'hétérogénéité de ses résultats et ses conséquences en termes d'équité et de redistribution ;
- la prise en compte des externalités et des effets de second tour.

²² Les articles d'Attanasio, Meghir et Santiago (2012) et de Low et Meghir (2017) sont des exemples de cette approche.

Chapitre 1

Les raisons et principes de l'application du calcul socioéconomique aux politiques d'investissement social

1. Du développement des mesures d'impact et du calcul socioéconomique

Les évaluations des investissements sociaux sont de nature et de qualité fortement variables. Elles prennent parfois la forme de descriptions évocatrices, voire anecdotiques, mais elles peuvent aussi provenir d'expérimentations aléatoires contrôlées. Les études qui les produisent concernent tout autant des comparaisons de pays que des individus. Elles reposent soit sur des statistiques simples mettant en évidence des tendances ou des écarts à la moyenne, ou bien sur des méthodes d'inférence causale plus rigoureuses²³. Cette diversité d'approches n'est pas pour autant satisfaisante. En l'absence de preuves empiriques robustes, résultant de modèles statistiques permettant d'identifier des causalités, il est à craindre que le concept d'investissement social ne se réduise à un « outil de stratégie politique, voire à un élément de rhétorique stratégique²⁴ ».

Toutefois, la situation a récemment évolué : **au cours des vingt dernières années, le nombre de travaux d'évaluation d'impact, expérimentaux ou quasi expérimentaux, a crû de façon exponentielle**. Ces méthodes sont aujourd'hui à la frontière de la recherche pour identifier, à partir des données, des résultats interprétables comme un lien de cause à effet. Elles utilisent pour cela des designs de recherche pour garantir que le fait de bénéficier de la politique étudiée n'est pas lié aux caractéristiques des personnes et qu'ainsi, la différence de résultat est seulement due à la politique. Ces méthodes visent à identifier statistiquement le contrefactuel – la situation sans la politique évaluée – et à le comparer avec les données observées. Pour y parvenir, la méthode expérimentale, qui consiste à tirer au sort les bénéficiaires de la politique évaluée et à comparer leur situation avec celle de ceux que le tirage au sort a exclus, garantit l'existence d'une relation de causalité entre la politique et l'effet observé, sans que l'analyste ait besoin de faire d'autres hypothèses. Les

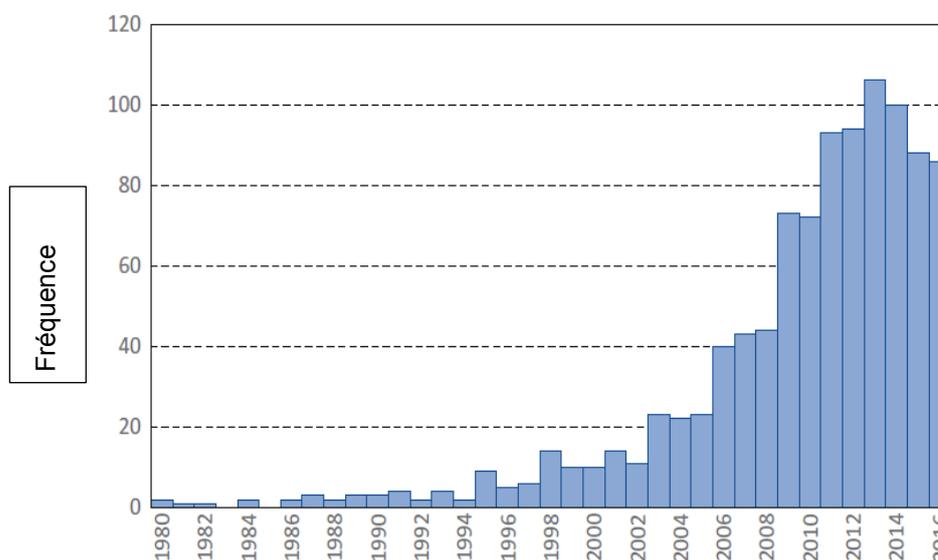
²³ Burgoon B. (2015), "Practical pluralism in the empirical study of social investment: Examples from active labour-market policy", dans *The Uses of Social Investment*, édité par A. Hemerijck, Oxford: Oxford University Press, p. 161-173.

²⁴ Damon J. (2015), « L'investissement social : contenu et portée d'une notion en vogue », *Revue de droit sanitaire et social*, n° 4, p. 722-733.

autres méthodes dites quasi expérimentales cherchent à identifier des situations où, conditionnellement à un certain nombre de facteurs, le fait de bénéficier de la politique publique est tout aussi indépendant des résultats que s'il avait été décidé au hasard²⁵. Ces méthodes reposent sur l'hypothèse d'indépendance conditionnelle et peuvent être regroupées en trois groupes :

- On compare bénéficiaires et non-bénéficiaires « toutes choses égales par ailleurs » en neutralisant les différences liées aux caractéristiques observables et on fait l'hypothèse qu'il n'y a pas de différences inobservables. On utilise pour cela des méthodes de régression ou d'appariement (*matching*).
- On utilise la trajectoire dans le temps d'un groupe non bénéficiaire comme contrefactuel en supposant qu'en l'absence d'intervention, celle du groupe des bénéficiaires aurait été parallèle à celle du groupe non bénéficiaire. Cette méthode exploite la dimension longitudinale ou pseudo-longitudinale²⁶ des données et fournit une interprétation causale si le biais qui caractérise les bénéficiaires est constant dans le temps (méthode des doubles différences).
- On trouve dans la « nature » une situation qui induit un changement dans la probabilité de bénéficier de la politique sans qu'il y ait un lien avec les résultats. Ces stratégies regroupent les régressions avec discontinuité et les variables instrumentales.

Figure 1– Nombre d'expérimentations aléatoires contrôlées réalisées entre 1980 et 2016 dans le domaine des politiques éducatives et ayant donné lieu à une publication dans une revue scientifique internationale



Source : d'après P. Connolly, C. Keenan et K. Urbanska (2018), "The trials of evidence-based practice in education: a systematic review of randomised controlled trials in education research 1980–2016", Vol. 60(3), *Educational Research*, p. 276-291

²⁵ Voir le guide Desplatz R. et Ferracci M. (2016), « Comment évaluer l'impact des politiques publiques ? Un guide à l'usage des décideurs et praticiens », Document de travail, France Stratégie.

²⁶ Les observations peuvent ne pas être celles des mêmes individus suivis mais de répétitions d'échantillons aléatoires d'une même population et former un « pseudo panel ».

Ces travaux se sont développés dans plusieurs champs de recherche, sans être pour autant directement liés à la notion d'investissement social. Par exemple, dans le domaine des politiques éducatives, le nombre d'expérimentations aléatoires contrôlées (« *randomized controlled trials* », ou RCT) ayant donné lieu à des publications internationales est passé de quelques unités en 1980 à plus de 80 par an depuis 2010 (Figure 1). Les évaluations quasi expérimentales ont suivi une tendance similaire et l'ensemble constitue aujourd'hui ce que certains appellent une « révolution empirique ». Ces travaux et les évaluations chiffrées qu'ils contiennent sont des ressources de première importance pour le choix, la conception et la mise en œuvre des politiques publiques d'investissement social.

Cependant, comparer et synthétiser les résultats de politiques voisines (par exemple, des programmes préscolaires destinés à la petite enfance) obtenus dans des périodes, des contextes ou des pays différents est un exercice difficile. La multiplication d'évaluations peut donc être un frein à leur utilisation par les autorités publiques. C'est pourquoi il convient de capitaliser sur les recherches existantes et fournir des indicateurs synthétiques agrégeant la connaissance passée.

En outre, les investissements sociaux peuvent affecter plusieurs aspects de la vie des individus (emploi, revenus, conjugalité, parentalité, santé, logement, etc.), parfois de façon complémentaire, parfois de façon contradictoire, ceci plus ou moins volontairement. Dès lors, il semble nécessaire d'intégrer les travaux d'évaluation d'impact dans un cadre d'analyse permettant de recenser et quantifier les effets des investissements sociaux en tenant compte de l'environnement social et institutionnel dans lequel ils s'inscrivent : **l'évaluation socioéconomique.**

L'analyse économique a développé un champ de recherche consacré à l'évaluation des politiques publiques afin d'en estimer les conséquences sur la société. De façon très simplifiée, **il existe deux types d'approches économiques** : l'une, plutôt normative et très formalisée, s'appuie sur des modèles économiques calibrés à partir de données empiriques ; l'autre, plus comptable, privilégie des exercices de simulation et/ou de projection pour estimer la valeur des différents coûts et bénéfices anticipés des politiques : **les analyses coûts-bénéfices.**

L'intérêt pour ces dernières n'a fait que croître au cours des décennies passées en France mais surtout à l'étranger, si bien qu'en 2010 une revue académique éditée par Cambridge University Press a vu le jour : *The Journal of Benefit-Cost Analysis*.

Historiquement, ces analyses ont été souvent utilisées pour évaluer le rendement socioéconomique de grands investissements publics, par exemple dans les secteurs des transports, de l'énergie, de l'environnement et du développement durable. De récentes évolutions administratives ont même élargi les exigences en matière d'évaluation socioéconomique des investissements publics. La loi de programmation pluriannuelle des finances publiques pour les années 2012 à 2017 stipule en effet en son article 17 que « les projets d'investissements civils financés par l'État, ses établissements publics, les établissements publics de santé ou les structures de coopération sanitaire font l'objet d'une évaluation socioéconomique préalable ». La loi précise de plus que « lorsque le montant

total du projet et la part de financement apportée par ces personnes²⁷ excèdent des seuils fixés par décret, cette évaluation est soumise à une contre-expertise indépendante préalable ». Une évaluation est donc obligatoire lorsque le projet dépasse 20 millions d'euros et une contre-expertise est nécessaire pour les projets de plus de 100 millions d'euros.

France Stratégie a largement contribué à définir et harmoniser ces pratiques en produisant plusieurs travaux, guides méthodologiques et évaluations dans ces domaines.

Deux travaux majeurs ont ainsi été rédigés :

- Quinet E. (2013), *L'évaluation socioéconomique des investissements publics*, France Stratégie ;
- France Stratégie et DG Trésor (2017), *Guide de l'évaluation socioéconomique des investissements publics*.

Ce dernier définit le calcul socioéconomique comme une méthode d'analyse dont le rôle est :

- d'objectiver les effets d'un investissement sur le bien-être de l'ensemble des agents affectés par cet investissement ;
- d'optimiser le coût de l'investissement au regard des objectifs qui lui ont été fixés en permettant une discussion objective des diverses options techniques et financières envisageables ;
- d'appréhender les risques et incertitudes entourant un investissement ;
- de fournir des éléments pour choisir entre des projets concurrents.

Néanmoins, jusqu'alors, les politiques sociales ou éducatives n'ont pas fait l'objet de développements spécifiques dans ces travaux. Comme cela a été précédemment indiqué, la différence essentielle provient du fait que les politiques d'investissement social concernent directement des personnes (enfants, élèves, étudiants, salariés, chômeurs, patients faisant l'objet de soins, etc.) alors que les grands investissements publics sont principalement relatifs à des travaux d'infrastructures (construction d'hôpitaux, de routes, de voies ferrées, etc.).

Le secteur de la santé fait exception puisqu'un champ assez fécond s'est développé autour d'évaluations coûts-efficacité et coûts-utilité (voir *infra*).

Il semble donc utile – pour ne pas dire indispensable – de s'intéresser à la mesure de l'efficacité des investissements sociaux. Certes, l'exercice n'est pas exempt de limites, voire de critiques, mais appliqué avec sérieux, en s'inspirant des principes généraux des analyses coûts-bénéfices, il permet de juger de l'efficacité et de l'équité à court, mais aussi à plus long terme, des politiques envisagées. Quelle que soit la politique considérée, l'analyse de son efficacité repose sur des principes et des éléments de méthode communs qui seront présentés dans un chapitre ultérieur.

²⁷ Ces « personnes » sont les financeurs publics cités précédemment, à savoir l'État, ses établissements publics, les établissements publics de santé ou les structures de coopération sanitaire.

Il faut néanmoins prendre garde et ne pas penser que l'efficacité d'une politique d'investissement social puisse se résumer à un chiffre unique, déduit d'un ensemble de données d'observation parfaitement adéquat. Il n'existe ni modèle ni chiffrage unique, pas plus qu'il n'existe une et une seule banque de données permettant de réaliser les exercices quantitatifs. **Une évaluation doit être adaptée à la politique d'investissement social considérée. Pour être crédible et utile, elle doit proposer des simulations et des analyses de sensibilité qui permettent de juger avec suffisamment de précision et de robustesse de son efficacité.**

Il faut également préciser que le calcul socioéconomique est un outil d'aide à la décision et ne doit pas être considéré comme un substitut au processus de décision démocratique, législatif ou administratif. Réciproquement, le processus de décision politique ne doit pas interférer dans la mise en œuvre technique et opérationnelle de l'analyse coûts-bénéfices.

Que retenir ?

Mesurer le rendement d'un investissement social implique de disposer d'au moins une évaluation rigoureuse de ses effets. L'existence d'évaluations d'impact est donc un prérequis.

Depuis trente ans, ces évaluations se sont multipliées et leur qualité scientifique n'a fait que croître, fournissant des sources de données abondantes qui peuvent être utilisées pour des analyses coûts-bénéfices.

En parallèle, le calcul socioéconomique s'est beaucoup développé, au point de devenir une obligation légale en France pour les investissements publics de plus de 20 millions d'euros. Cependant, son application aux politiques sociales est restée limitée, à l'exception des politiques de santé qui ont largement recours aux analyses coûts-efficacité.

Il est possible et même souhaitable d'utiliser les méthodes d'évaluation socioéconomique enrichies par les techniques, les données et les résultats d'évaluations d'impact afin d'estimer le rendement d'investissement sociaux.

Une telle démarche ne doit toutefois pas se substituer à la décision publique ni même dépasser le cadre d'un outil d'aide à la décision.

2. Qu'est-ce qui distingue le calcul socioéconomique appliqué à l'investissement social des autres formes d'évaluation ?

Le calcul socioéconomique cherche à exprimer en termes monétaires la valeur sociale des effets d'une politique publique en les comparant aux coûts de sa mise en œuvre. L'objectif premier de l'analyse coûts-bénéfices, telle que définie par le guide de France Stratégie (2017), est « d'objectiver les effets d'un investissement sur le bien-être de l'ensemble des agents affectés par cet investissement afin d'évaluer sa capacité à répondre à un besoin identifié, en appréciant les gains attendus au regard des coûts engendrés. Compte tenu de

tous ces effets, l'objectif premier de l'évaluation socioéconomique est de fournir une analyse permettant de prendre une décision sur l'opportunité de réaliser un projet ».

C'est un cadre d'analyse complet visant à tenir compte de tous les bénéfices et coûts sociaux de façon explicite et détaillée. Ayant été jusqu'alors principalement employée dans d'autres domaines (équipements publics, environnement, etc.), l'analyse coûts-bénéfices est de plus en plus appliquée dans le champ social, éducatif ou en santé²⁸. Cette méthode aide à identifier, parmi un ensemble de politiques possibles, la plus efficiente, c'est-à-dire celle qui produit les bénéfices les plus importants à partir des ressources disponibles. L'analyse coûts-bénéfices a toutefois des limites : elle ne permet pas de juger de la désirabilité sociale d'une politique publique²⁹ ou des enjeux de redistribution qui lui sont associés. Elle se focalise sur la maximisation de ses rendements monétisables. Or les considérations en termes de redistribution, de justice sociale et d'éthique sont souvent les motivations premières de nombreuses politiques sociales. De ce fait, les méthodes habituelles d'analyse coûts-bénéfices ne peuvent être appliquées directement et sans précaution aux politiques sociales.

Dans le cinquième chapitre, nous décrivons plus en détail les méthodes permettant de tenir compte des enjeux éthiques et redistributifs. De façon générale, leur prise en compte suppose un niveau d'information qui va beaucoup plus loin que la simple répartition des coûts et des effets entre des sous-populations ou parties prenantes.

Le besoin de tenir compte de l'hétérogénéité des effets individuels, les questions de justice sociale liées aux inégalités d'origine ou de destin, etc., font de l'application du calcul socioéconomique aux investissements sociaux un exercice difficile. Ces problèmes justifient que l'on doive de plus en plus souvent mobiliser une méthode jusqu'alors moins usitée, la microsimulation. Comme le rappellent Blanchet *et al.*³⁰ (2016), les modèles de microsimulation permettent en théorie de tout simuler, des mécanismes produisant tout à la fois des effets individuels et collectifs, mais aussi des effets de pairs ou de dissémination, etc. La contrepartie de cette flexibilité est la lourdeur des modèles sous-jacents, mais aussi le besoin de données très riches et d'hypothèses nombreuses, avec de ce fait un « risque modélisateur » avec possibilité d'incohérences. Même en cas de spécification parfaite du modèle, le risque est que la complexification de la microsimulation se fasse au détriment de sa transparence et de sa lisibilité. Le bien-fondé de l'utilisation de cette méthode a été récemment reconnu par Heckman et ses coauteurs (2017)³¹ qui assimilent à un exercice de

²⁸ Voir notamment Weimer D. et Vining A. (2009), *Investing in the Disadvantaged: Assessing the Benefits and Costs of Social Policies*, Georgetown University Press, 304 p.

²⁹ La désirabilité sociale doit être intégrée en amont du choix des interventions à analyser. Par exemple, l'analyse socioéconomique peut être appliquée à une politique de prévention de la sécurité routière par l'installation de radars automatiques mais ne dit rien quant au souhait de la population pour cette politique.

³⁰ Blanchet D., Hagneré C., Legendre F. et Thibault F. (2016), « Évaluation des politiques publiques, *ex post* et *ex ante* : L'apport de la microsimulation », *Revue économique*, Vol. 67(4), p. 685-696. Didier Blanchet a exposé devant le groupe de travail le contenu de cet article, ainsi que d'autres considérations sur l'utilité de la microsimulation.

³¹ García J. L., Heckman J., Leaf D. E. et Prados M. J. (2017), "Quantifying the Life-cycle Benefits of a Prototypical Early Childhood Program", *NBER Working Paper No 23479*. Ils écrivent notamment : "Our microsimulation model starts with information on observed characteristics at age 30, along with the information on observed characteristics available at this age" (Garcia *et al.*, 2017, p. 32).

microsimulation leur analyse coûts-bénéfices du programme préscolaire *Abecedarian* mis en œuvre aux États-Unis. Pour prendre en compte les considérations éthiques et redistributives d'une politique publique, il est également possible de recourir aux méthodes de l'économie comportementale ou expérimentale. L'objectif est alors de faire révéler les préférences pour la redistribution et de les intégrer soit sous forme de pondérations, soit d'utiliser des valeurs sociales des effets monétisés qui intègrent déjà les préférences pour la redistribution³².

Dans le cas particulier des politiques sociales, il est par ailleurs difficile de conduire une analyse coûts-bénéfices suffisamment complète, notamment parce que l'on dispose rarement d'enquêtes de suivi sur longue période (dix ou vingt ans, voire plus). Pour le dire autrement, les données longitudinales manquent bien souvent. Lorsqu'elles existent, elles couvrent au mieux les trois à cinq années qui suivent la mise en place de la politique examinée. Il est pourtant crucial de mener une analyse aussi longue que possible car les politiques d'investissement social sont par définition censées produire des effets significatifs souvent après plusieurs années, dans différentes sphères de la vie publique et privée. Par exemple, une politique d'accueil des jeunes enfants est susceptible d'affecter l'employabilité des parents, et notamment des mères, mais aussi d'agir sur le développement cognitif et non cognitif de ces enfants, d'améliorer leur réussite scolaire ultérieure, leur insertion dans la vie active, mais aussi leur comportement civique. Limiter en ce cas l'analyse aux seuls effets sur l'emploi des parents ou aux seuls effets sur la cognition des enfants distordrait de façon importante l'évaluation du rendement social de cette politique.

Enfin, l'observation des effets de long terme ne suffit pas ; encore faut-il disposer de méthodes rigoureuses permettant de voir si l'on peut interpréter les corrélations estimées comme des liens de cause à effet. À la différence des grands investissements publics où les évaluations *ex-post* sont moins nombreuses, l'investissement social bénéficie aujourd'hui d'un ensemble de travaux assez conséquent pouvant être mobilisé pour calibrer les analyses coûts-bénéfices, notamment celles que l'on peut conduire *ex ante* alors que tous les effets escomptés ne sont pas encore survenus. L'identification de la direction et de l'intensité de la relation causale est aujourd'hui au cœur de l'analyse des politiques sociales et éducatives, et les méthodes statistiques qui y sont employées se distinguent de celles utilisées dans d'autres domaines de recherche, par exemple dans des études de nature plus macroéconomique.

Il faut alors prendre garde au fait que beaucoup de ces politiques d'investissement social, par exemple celles destinées à la petite enfance, ont été mises en œuvre dans des pays ou dans des contextes institutionnels très spécifiques, notamment ceux prévalant aux États-Unis, et donc qu'il est risqué de transposer sans précautions leur conception et leurs résultats directement au cas français.

Les résultats de recherches menées dans d'autres pays et/ou dans des périodes différentes sont difficiles à extrapoler. Les conditions pour y parvenir ne sont pas évidentes et elles ont rarement une portée générale. Mais certaines équipes de recherche font le pas et définissent

³² Par exemple, si l'on dispose d'évaluations contingentes dans lesquelles la population indique son consentement à payer pour éviter qu'un enfant décroche, en distinguant le consentement à payer pour les enfants issus de quartiers prioritaires des autres, ces valeurs intègrent les préférences pour la redistribution et peuvent être utilisées comme telles dans l'analyse coûts-bénéfices.

un cadre pour obtenir des résultats « généralisables ». C'est le cas de J-Pal³³ qui définit son « *generalisability framework* » en fonction de quatre étapes³⁴ :

- étape 1 : définir le modèle théorique sous-jacent : quels mécanismes sont susceptibles d'être responsables des effets observés ?
- étape 2 : les conditions locales permettent-elles l'application de ce modèle ?
- étape 3 : dispose-t-on de preuves suffisantes pour affirmer que le changement de comportement postulé provoque un effet significatif ?
- étape 4 : quelles sont les preuves dont on dispose pour affirmer que le projet pourrait être mis en œuvre dans un autre contexte ? Dit autrement, est-on suffisamment certain que les résultats antérieurs ne dépendaient pas trop de l'équipe opérationnelle, de la localité, de la population cible, etc. ?

Une autre façon de vérifier la validité des recherches passées est de recenser les résultats de façon systématique et de les synthétiser dans une méta-analyse (voir chapitre 2).

Que retenir ?

Le calcul socioéconomique cherche à mesurer la valeur sociale des effets d'une politique publique en les comparant aux coûts de sa mise en œuvre.

C'est un cadre d'analyse complet visant à tenir compte de tous les bénéfices et coûts sociaux de façon explicite et détaillée, mais qui doit être adapté pour pouvoir être appliqué aux investissements sociaux.

L'investissement social se distingue parce qu'il concerne principalement des personnes, qu'il se dote d'objectifs ayant directement trait aux enjeux d'équité et de redistribution, que ses effets sont souvent diffus et hétérogènes suivant les populations, que ses objectifs peuvent bénéficier à certains groupes mais pas à d'autres, etc.

La prise en compte de ces spécificités impose d'avoir recours à un cadre d'analyse et à des méthodes spécifiques, comme les outils de microsimulation, et de disposer de données individuelles et de suivi longitudinaux sur longues périodes.

³³ J-PAL, Laboratoire d'Action contre la Pauvreté, est un réseau de 131 chercheurs affiliés dans 40 universités du monde entier dont la mission consiste à lutter contre la pauvreté en veillant à ce que les politiques sociales s'appuient sur des preuves scientifiques. Voir le site <https://www.povertyactionlab.org/fr>

³⁴ Bates M. A. et Glennerster R. (2017), "The Generalizability Puzzle", *Stanford social innovation review*, summer, p. 50-54.

3. Dans quelle phase du processus évaluatif doit-on tenir compte des bénéfices et des coûts anticipés d'une politique d'investissement social ?

Le calcul socioéconomique peut être mobilisé en amont de la mise en œuvre d'une politique particulière, lorsqu'il faut par exemple décider du choix de ses modalités ou arbitrer entre plusieurs décisions. Il s'agit de l'évaluation dite *ex ante*. Il peut l'être également pendant la phase de mise en place de cette politique, ou bien *a posteriori*, pour juger de l'efficacité globale de l'investissement réalisé et pour orienter de futures décisions. Il s'agit de l'évaluation dite *ex post*.

La réalisation d'une analyse coûts-bénéfices n'a pas la même temporalité que le déploiement de la politique devant être évaluée. La production d'une analyse complète et rigoureuse est un processus souvent plus long que la mise en œuvre d'une politique publique. **Cette différence est souvent source de tension entre l'expert et le décideur**, le premier ne pouvant fournir des estimations de qualité à temps, le second devant faire des choix à partir de calculs trop exploratoires.

C'est pourquoi il serait souhaitable de disposer d'un registre des politiques d'investissement social déjà réalisées, pour lesquelles il existe une littérature scientifique suffisamment abondante ayant produit des évaluations de qualité : parce que ce registre permet d'estimer le rendement social de chacune des politiques recensées, il peut être utilisé dans une logique prospective et comparative.

C'est notamment ce que propose le Washington State Institute for Public Policy³⁵ en mettant à la disposition du public et des autorités plusieurs dizaines d'analyses coûts-bénéfices de politiques sociales. **Ses calculs reposent la plupart du temps sur des méta-analyses, qui peuvent être rédigées lorsqu'existent plusieurs travaux d'évaluation relatifs à la même intervention ou à une catégorie d'interventions comparables.** Toutes les évaluations utilisent un cadre méthodologique commun qui permet de comparer les rendements de politiques alternatives.

Même s'il est difficile de réaliser une évaluation *ex ante* de qualité avant la mise en œuvre d'une politique, il ne faut pas pour autant y renoncer. Bien au contraire, il est souhaitable de conduire une analyse socioéconomique en amont du projet, de façon à anticiper ses effets possibles à partir des données préexistantes et/ou de travaux menés dans d'autres contextes ou d'autres pays. **Cette analyse préalable fournit un cadre de référence qui peut être ensuite comparé aux évaluations conduites pendant et après la mise en œuvre du programme.**

Une vertu de cette approche est d'amener les parties prenantes à adopter une démarche réflexive, c'est-à-dire à s'interroger sur les effets possibles de la politique, à adopter un raisonnement contrefactuel, finalement, à se poser des questions qu'ils ne se poseraient pas autrement. Lorsque les données sur la situation des participants au programme sont

³⁵ Voir [le site internet du WSIPP](#). Le chapitre 4 contient une présentation détaillée d'une analyse socioéconomique produite par cet institut.

recueillies de manière régulière et suffisamment répétée au cours du temps, il est possible de réviser au fur et à mesure les conclusions des analyses coûts-bénéfices préalables, de façon à produire une évaluation plus précise du programme. Cet exercice peut permettre de réviser certaines modalités de la politique évaluée, ou, dans certains cas, de l'abandonner au profit d'une politique plus adéquate. Dans ces conditions, l'analyse coûts-bénéfices *ex ante* sert de « benchmark ». C'est un outil important d'aide à la décision renseignant sur l'ampleur des effets attendus, les facteurs de risques et, par ailleurs, elle aide à concevoir les besoins pour l'évaluation *ex post* et permet *in fine* une meilleure lisibilité de ces résultats.

Pour que la révision régulière soit possible, **l'évaluation socioéconomique ex post doit être conçue avant même la mise en œuvre de la politique** de façon à prévoir le recueil des données indispensables et à pouvoir réaliser des comparaisons valides avec d'autres politiques. L'objectif est de disposer d'une évaluation de préférence quantitative – par exemple fondée sur une expérimentation aléatoire contrôlée – permettant de mesurer les effets de l'investissement social sur les variables qu'il est censé affecter (le niveau d'éducation, le revenu, l'état de santé, le recours à l'aide sociale, etc.). Cependant, beaucoup de politiques d'investissement social peuvent ne pas être adaptées à la mise en œuvre d'une expérience aléatoire pure. Mais ce n'est pas pour autant qu'elles doivent être d'emblée écartées.

Des protocoles de recherche ingénieux permettent de lever certaines contraintes de mise en œuvre tout en autorisant l'identification précise des causalités à l'œuvre. Par exemple, dans le cas où l'investissement est de grande ampleur, il est possible de tirer profit d'une mise en œuvre progressive ou différée dans le temps (*phase-in design*), comme c'est le cas pour le programme Parler Bambin³⁶. Si la politique est un droit pour lequel il n'est pas possible de contraindre ou d'interdire aux individus d'en bénéficier, encourager au hasard des personnes éligibles peut fournir une source d'identification de l'impact. C'est ce design qui a été mis en place pour l'évaluation du projet d'investissement social Reliance³⁷, dispositif emblématique de la politique d'investissement social de la Cnaf³⁸.

³⁶ Il s'agit d'un programme expérimental de formation des personnels des établissements d'accueil des jeunes enfants dont le déploiement s'étale sur six ans ; cela permet de comparer les enfants affectés aux premières crèches bénéficiaires du programme aux enfants relevant de celles non encore entrées dans le dispositif. Voir <http://www.parlerbambin.fr/le-programme-national/article/le-programme-de-recherche>

³⁷ Il s'agit d'un programme expérimental d'accompagnement global à la reprise d'activité durable à destination des chef.fe.s de famille monoparentale bénéficiaire du RSA depuis au moins deux ans. Le programme d'évaluation piloté par Arthur Heim s'appuie sur une convocation aléatoire à participer à des réunions d'information qui visent à inciter les personnes à participer au projet. Voir le site http://www.arelia-asso.fr/index.php?option=com_content&view=article&id=510:reliance-est-un-nouveau-dispositif-experimental&catid=25:newsletterlirelasuite

³⁸ Caisse nationale des allocations familiales.

Que retenir ?

La réalisation d'une analyse coûts-bénéfices n'a pas la même temporalité que le déploiement de la politique devant être évaluée, ce qui est souvent source de tension.

Afin d'être en capacité de produire une évaluation rigoureuse ex ante avant le lancement du projet, il est souhaitable de disposer d'un registre d'analyses coûts-bénéfices ex ante.

Les décalages temporels ne doivent pas conduire à négliger les évaluations ex ante. Au contraire, une analyse coûts-bénéfices ex ante apparaît comme un outil primordial d'aide à la décision, conduit les parties prenantes à s'engager dans une démarche réflexive et sert de « benchmark » pour l'évaluation ex post.

L'évaluation ex post doit être conçue en amont et en lien direct avec l'évaluation ex ante.

Il faut organiser autant que possible des évaluations expérimentales et ne pas hésiter à profiter des contraintes de mise en œuvre des politiques pour instaurer des protocoles d'expériences adaptés.

Chapitre 2

Définition et déroulé de l'évaluation d'une politique d'investissement social

L'application du calcul socioéconomique à l'évaluation d'une politique d'investissement social suppose de suivre une série d'étapes indispensables au cours desquelles plusieurs choix méthodologiques doivent être faits. Elle requiert une définition claire des objectifs de la politique mais aussi des informations suffisantes sur les coûts encourus et les bénéfices attendus. Ce chapitre décrit les étapes et éléments nécessaires à la conduite d'une analyse coûts-bénéfices.

1. Cadrage général

Les objectifs des politiques d'investissement social doivent être clairement définis. Ces objectifs s'attachent généralement à résoudre un problème social ou à offrir une intervention. Le problème, les objectifs et les solutions doivent être clairement énoncés mais sans préjuger de la manière d'agir, la gamme des interventions possibles devant rester assez ouverte. Une fois les objectifs définis, le calcul socioéconomique peut être envisagé : il aide à choisir le type d'intervention en évaluant des alternatives ou amendements, mais aussi à piloter la mise en œuvre concrète en fournissant des mesures du rendement social à différents moments.

Un préalable à l'analyse socioéconomique consiste à évaluer les investissements sociaux envisagés à différents égards³⁹ :

- la **pertinence** : la politique envisagée permet-elle de répondre aux objectifs globaux, répond-elle aux besoins identifiés ?
- la **cohérence** : La politique envisagée est-elle cohérente vis-à-vis d'autres politiques, des moyens matériels et humains consacrés ou disponibles à court terme, des contraintes financières plus générales, des objectifs ? Pour illustrer des enjeux potentiels d'incohérence, on pourrait signaler à titre d'exemple la coexistence d'objectifs de mixité scolaire et d'un secteur privé subventionné « sous contrat » sélectionnant ses élèves. Ou

³⁹ Méthodes d'évaluation des politiques publiques : actes du séminaire, sous la direction de N. Studer (2011), DREES, Paris.

encore la conjugaison d'actions de prévention santé avec des publicités dans les médias publics concernant des aliments gras, salés ou sucrés, etc.

Par ailleurs, les objectifs de l'investissement social répondent souvent à des besoins spécifiques qui concernent des groupes de personnes qu'il faut identifier et caractériser au préalable. Quels sont les bénéficiaires pressentis ? Vise-t-on une population spécifique ou plus générale, s'agit-il d'objectifs ciblés sur un ou plusieurs territoires ? Avec qui mettre en œuvre la politique choisie ? Quels sont les acteurs susceptibles d'être impliqués dans la mise en œuvre et la gestion de la politique, quelles seront leurs fonctions ?

La temporalité de la mise en œuvre doit également être caractérisée à la fois :

- par sa dimension opérationnelle : combien de temps faut-il pour mettre en place la politique choisie ? Doit-elle être déployée progressivement ou en une seule phase ?
- et par son processus d'évaluation : combien de temps faut-il pour que les effets attendus puissent éventuellement apparaître ? Ces effets sont-ils pérennes ? À quel horizon ?

En outre, il est indispensable de réaliser au préalable une recension de la littérature internationale afin de savoir si des politiques similaires ont été mises en œuvre dans d'autres pays, et s'il existe des évaluations de leurs effets. La relative abondance des travaux internationaux préexistants peut conditionner le type d'évaluation qu'il convient de mener, mais aussi permettre de dresser une première liste des bénéfices et des coûts possibles. La dernière section de ce chapitre se consacre à cet enjeu crucial.

Certes, l'évaluation socioéconomique *ex ante* ne peut s'effectuer pleinement que lorsque le projet a atteint un niveau de maturité relativement avancé. Il importe toutefois de conduire des études préliminaires : l'évaluation socioéconomique ne doit pas être effectuée trop en aval de la définition et de la mise en œuvre de la politique choisie. L'évaluation *ex ante*, qui peut par exemple résulter d'une expérimentation pilote, a ainsi pour objectif d'éclairer l'opportunité d'une réforme, le choix entre les différentes options envisagées, elle doit intervenir avant que les décisions relatives à la mise en œuvre de la politique ne soient prises.

Avant de débiter une analyse socioéconomique, il importe de caractériser de façon précise la situation existante et ses évolutions possibles dans un contexte donné (scénario de référence), indépendamment de la politique envisagée. En effet, le cœur de l'analyse socioéconomique réside dans la comparaison à *la marge* des bénéfices et des coûts par rapport à la situation contrefactuelle, c'est-à-dire celle qui prévaudrait en l'absence d'intervention. Il est donc crucial que cette situation contrefactuelle soit correctement décrite et analysée afin de pouvoir identifier, quantifier et valoriser les éléments spécifiques constitutifs de la réforme ou de la nouvelle politique.

Cette caractérisation peut être assez littérale ou plus quantitative, notamment lorsque l'investissement social envisagé implique une modification des ressources employées⁴⁰.

⁴⁰ Des politiques d'investissement social peuvent aussi consister en un changement de pratiques, ou l'édiction de nouvelles règles qui ne modifient pas les ressources allouées mais changent les pratiques. Dans ce cas, le coût marginal peut être très faible.

Prenons deux exemples pour illustrer ce propos. La stratégie de lutte contre la pauvreté, récemment annoncée par le gouvernement, s'inscrit largement dans une logique d'investissement social en se proposant de lutter contre les causes de la pauvreté dès les premiers âges de la vie. L'un des objectifs est ainsi de renforcer les compétences des enfants d'origines sociales défavorisées. Pour répondre à cet objectif, plusieurs mesures vont être mises en œuvre, comme par exemple une réforme de la formation des personnels des établissements d'accueil des jeunes enfants (EAJE) et des assistant(e)s maternel(le)s, et l'augmentation des taux d'encadrement dans les établissements scolaires appartenant au réseau d'éducation prioritaire.

Le calcul socioéconomique se propose d'estimer, à l'aide d'une même méthode, le rendement social de ces différentes mesures afin de comparer leur niveau d'efficacité, et ce par rapport à une situation de référence bien définie. Dans le premier cas, il faut donc caractériser le niveau de formation des personnes concernées, détailler les forces et les faiblesses de la formation continue qui leur est pour l'instant proposée, décrire les familles accueillies dans les différents modes de garde, etc., et recueillir des éléments sur les coûts des dispositifs actuels. Dit autrement, il faut pouvoir décrire de façon précise ce que serait la situation sans intervention et, par là, mettre en valeur les changements opérés. Pour l'autre volet de ce plan, il faut disposer d'informations précises sur les taux d'encadrement en éducation prioritaire, le niveau des élèves, le profil et la rémunération des enseignants, etc. L'existence de substituts proches à certaines politiques peut changer la définition du contrefactuel et affecter l'identification des effets de l'investissement social. Ce point est discuté plus en détail dans la section 0 et illustré par un exemple empirique.

Que retenir ?

Le calcul socioéconomique appliqué à l'investissement social s'intègre dans un processus évaluatif plus général et ne doit pas s'y substituer. En particulier, les autres registres d'évaluation, qui concernent notamment la pertinence et la cohérence de l'intervention, ne doivent pas être ignorés.

Les objectifs de l'investissement social doivent être clairement définis de façon à identifier en amont les parties prenantes (bénéficiaires, non bénéficiaires, etc.), la temporalité du projet et ses effets potentiels.

Il est important de recenser et de connaître les recherches qui se rapportent à l'efficacité de politiques similaires afin de mieux en anticiper les effets.

Dans le cadrage général, l'analyste doit être en mesure de définir et de caractériser précisément la situation contrefactuelle, c'est-à-dire celle qui prévaudrait en l'absence d'intervention.

2. Adapter les méthodes d'évaluation aux questions posées

Les questions relatives à l'évaluation du rendement de l'investissement social doivent alors être formulées de façon claire. En premier lieu, **faut-il mettre en place une analyse coûts-efficacité ou coûts-bénéfices ?**

La réponse à cette question dépend de ce que l'on cherche à savoir. De façon très schématique, une analyse coûts-efficacité est particulièrement adaptée lorsqu'il s'agit de comparer plusieurs politiques cherchant à atteindre un même objectif. Cela suppose qu'il existe des études permettant d'identifier rigoureusement les effets respectifs de ces politiques, et proposant des mesures de leur efficacité à l'aide d'analyses *ex post*. Néanmoins, il est rarement possible de réaliser simultanément plusieurs analyses coûts-efficacité, à moins de disposer d'un registre d'évaluations d'impact et de pouvoir mesurer les coûts des politiques alternatives.

Par ailleurs, **les décideurs politiques ne mettent souvent en place qu'une seule politique dont les objectifs sont multiples**⁴¹. **Dans ces conditions, une analyse coûts-bénéfices (ACB ci-après) apparaît plus adaptée.**⁴² Cependant, les résultats de ces analyses restent soumis à un certain degré d'incertitude et leur utilisation dans le débat public doit être assortie de beaucoup de précautions.

Une question corollaire mais fondamentale est celle de l'évaluation *ex post* d'une politique d'investissement social. Dans la section 1.3, nous avons discuté de la réflexivité permise par l'articulation d'une évaluation *ex ante* et *ex-post*. Si la première permet de guider la décision, seule l'évaluation *ex post* peut identifier les impacts et *a fortiori* le rendement de l'investissement social considéré et confirmer ou non les résultats.

S'interroger, au moment même de la conception de la politique envisagée, sur la démarche d'évaluation, sur son protocole, sur sa réception par les acteurs susceptibles de la mettre en œuvre, est le meilleur moyen d'adapter les méthodes de calcul aux questions d'évaluation et aux contraintes de faisabilité. Il est également possible que la mise en œuvre d'une politique sociale fournisse des opportunités d'évaluation quasi expérimentale, par exemple lorsqu'elle est déployée progressivement sur le territoire, lorsque la sélection des bénéficiaires est soumise à un aléa administratif, etc. Lorsqu'il y a convergence des évaluations *ex ante* et *ex post*, cela valide le modèle *ex ante* et accroît son utilité pour de futures utilisations.

⁴¹ Deux exemples parmi d'autres : 1) le Crédit d'impôt pour la compétitivité et l'emploi (CICE) instauré par le gouvernement de Jean-Marc Ayrault a pour objectif « de redonner aux entreprises des marges de manœuvre pour investir, prospecter de nouveaux marchés, innover, favoriser la recherche et l'innovation, recruter, restaurer leur fonds de roulement ou accompagner la transition écologique et énergétique grâce à une baisse du coût du travail » ; 2) le dédoublement des classes de CP et de CE1 dans l'éducation prioritaire mis en place à la rentrée scolaire 2017 par le gouvernement d'Edouard Philippe a pour objectif global « 100 % de réussite en CP », soit « garantir, pour chaque élève, l'acquisition des savoirs fondamentaux - lire, écrire, compter, respecter autrui ».

⁴² La section 3.1 ci-après est consacrée à la présentation des deux méthodes, ACB et analyse coûts-efficacité.

Que retenir ?

L'analyse coûts-bénéfices et l'analyse coûts-efficacité ont chacune des avantages et sont plus ou moins adaptées à certaines questions de recherche ou politiques.

Certaines politiques ciblent un seul objectif ; en ce cas, une analyse coûts-efficacité apparaîtra plus judicieuse.

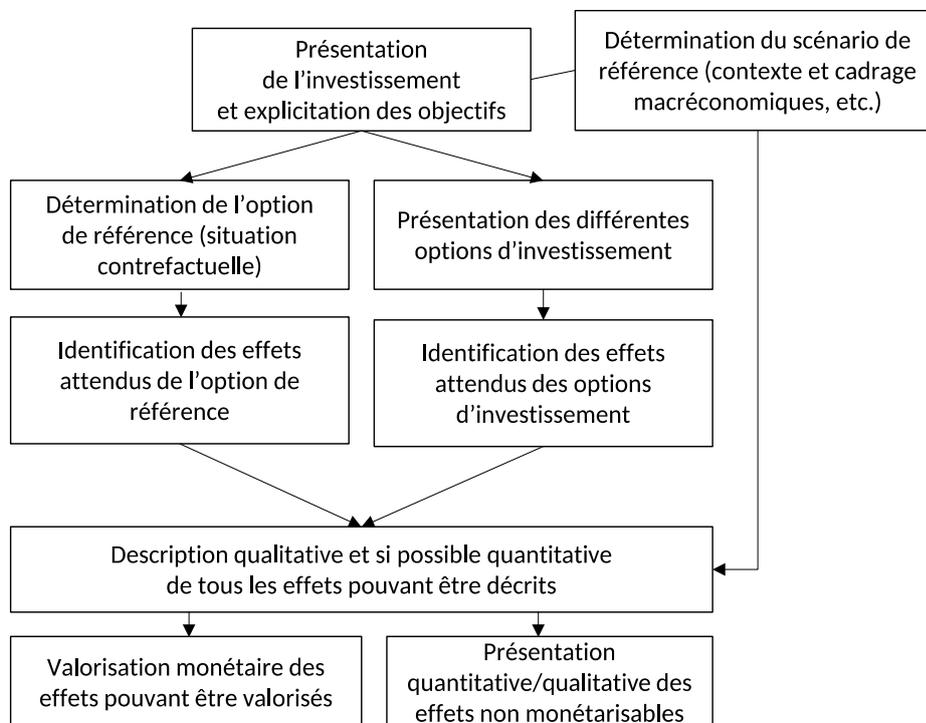
D'autres au contraire visent plusieurs objectifs, ou sont susceptibles d'engendrer des effets indirects. Dans ce cas, une analyse coûts-bénéfices est plus à même de rendre compte des impacts multiples de ces politiques.

Le lien entre les questions évaluatives et les méthodes invite une nouvelle fois à penser l'évaluation ex post en amont de la mise en œuvre des politiques d'investissement social.

3. Les étapes de la mise en œuvre du calcul socioéconomique

Conduire une analyse coûts-bénéfices consiste en une série d'étapes successives qui coïncident plus ou moins avec celles qui séquent la conception et la mise en œuvre d'une politique publique. Le guide méthodologique de France Stratégie et de la DG Trésor définit ces étapes qui sont résumées dans la Figure 2.

Figure 2 – Schéma des étapes systématiques de l'évaluation socioéconomique



Source : d'après France Stratégie et DG Trésor (2017)

Un premier bloc d'étapes consiste à définir de façon précise le cadre général en présentant l'investissement, ses objectifs et le scénario de référence. Dans un second temps, il faut définir les options d'investissement et la situation contrefactuelle puis les effets attendus dans les deux situations (intervention et contrefactuelle). Ensuite, les effets doivent être recensés, décrits et quantifiés autant que possible. Ceux qui peuvent être monétisés vont l'être dans l'étape suivante. Enfin, il faut construire le modèle coûts-bénéfices après avoir actualisé les effets et, pour finir, procéder à des analyses de sensibilité.

Ces étapes apparaissent très adaptées pour guider la mise en place d'une évaluation socioéconomique pour un projet d'investissement physique. Néanmoins, dans le cadre de l'investissement social, il nous semble nécessaire d'insister sur certaines étapes. En effet, nous voulons rendre explicite la présentation des différents effets et bénéfices distingués pour chaque partie prenante pour ne pas manquer la prise en compte des effets redistributifs et des enjeux d'équité. Il est également important de rendre visible la projection de ces effets et coûts dans le temps.

Les autres guides proposés par Roijnes et Renes (2013) du CPB Netherlands Bureau for Economic Policy Analysis, ou ceux du Treasury of New-Zealand Gouvernement (2015) sont sensiblement différents, comme le montre le Tableau 1.

Conformément aux besoins spécifiques de l'investissement social, nous retenons les étapes définies par Cellini et Kee (2015)⁴³ dans un chapitre consacré aux calculs socioéconomiques appliqués aux politiques sociales :

1. *Définir le cadre général d'analyse*
2. *Décider des coûts et bénéfices à imputer aux différents types d'acteurs concernés par l'analyse*
3. *Identifier et caractériser ces coûts et bénéfices*
4. *Prévoir l'évolution des coûts et des bénéfices tout au long de la mise en œuvre de la politique*
5. *Évaluer les coûts de l'investissement social*
6. *Quantifier les bénéfices par le biais d'une mesure de l'efficacité (pour une analyse coûts-efficacité) ou monétiser les bénéfices (pour une analyse coûts-bénéfices)*
7. *Actualiser les coûts et bénéfices futurs au moyen d'une valeur actualisée*
8. *Calculer un ratio coûts-efficacité (pour une analyse coûts-efficacité) ou une valeur actualisée nette socioéconomique (pour une analyse coûts-bénéfices)*
9. *Conduire une analyse de sensibilité*
10. *Faire des recommandations appropriées*

⁴³ Cellini S. R. et Kee J. E. (2015), "Cost effectiveness and cost benefit analysis", dans *Handbook of Practical Program Evaluation*, édité par K. Newcomer, H. Hatry et J. Wholey, 4^e édition, Jossey-Bass, p. 493-530.

Tableau 1 – Comparaison des étapes d'une analyse coûts-bénéfices dans différents guides méthodologiques internationaux

Guide de France Stratégie et DG Trésor (2015)	Cellini et Kee (2015)	Treasury of New-Zealand (2015)	Rojnes et Renes (2013)
Cadre général de l'analyse	Cadre général de l'analyse	Définir la politique et le contrefactuel	Analyse du problème
Recensement des différents effets des options d'investissement	Décider quels coûts et bénéfices de quelles parties prenantes intégrer dans l'analyse	Identifier les gagnants et les perdants de la politique	Définition de la situation contrefactuelle
Objectivation des effets non marchands, analyse qualitative et quantitative	Identifier et caractériser ces coûts et bénéfices	Identifier les coûts et bénéfices	Définir les alternatives de politique publique
Valorisation monétaire des effets de l'investissement	Prévoir l'évolution des coûts et des bénéfices tout au long de la mise en œuvre de la politique		Déterminer les effets et les bénéfices
Analyse coûts-bénéfices portant sur les effets monétarisés de l'investissement	Évaluer les coûts de l'investissement social	Monétiser les coûts et bénéfices	Déterminer les coûts
Analyser les risques et les incertitudes entourant les résultats de l'évaluation socioéconomique	Quantifier les bénéfices par le biais d'une mesure de l'efficacité (pour une analyse coûts-efficacité) ou monétiser les bénéfices (pour une analyse coûts-bénéfices)	Actualiser et comparer les coûts et les bénéfices	Analyser le risque et l'incertitude
Prendre en compte les effets non monétisables des investissements	Actualiser les coûts et bénéfices futurs au moyen d'une valeur actualisée	Évaluer la qualité de l'analyse coûts-bénéfices : d'autres recherches sont-elles nécessaires ?	Résumer l'analyse coûts-bénéfices
Présenter les résultats	Calculer un ratio coûts-efficacité (pour une analyse coûts-efficacité) ou une valeur actualisée nette socioéconomique (pour une analyse coûts-bénéfices)	Préparer le rapport final	Présenter les résultats
	Conduire une analyse de sensibilité		
	Faire des recommandations appropriées		

Source : France Stratégie et DG Trésor (2017), Guide de l'évaluation socioéconomique des investissements publics. Cellini S. R. et Kee J. E. (2015), "Cost effectiveness and cost benefit analysis", dans Handbook of Practical Program Evaluation, édité par K. Newcomer, H. Hatry et J. Wholey, 4^e édition, Jossey-Bass, p. 493-530. Romijn G. et Renes G. (2013), General Guidance for Cost-Benefit Analysis, CPB/PBL, Den Haag. Treasury of New Zealand Government (2015), Guide to social costs-benefits analysis.

L'ordre de ces étapes est assez logique ; toutefois, certaines peuvent être conduites en parallèle plutôt que séquentiellement. Des retours vers des étapes antérieures sont parfois nécessaires, notamment lorsque les circonstances ou le contexte économique changent.

3.1. Définir le cadre général d'analyse

La définition du cadre général d'analyse correspond à ce qui a été décrit en détail dans la section 1 de ce chapitre. En résumé, il s'agit d'identifier le problème et de définir les objectifs de la politique d'investissement social, d'interroger la pertinence et la cohérence de cette politique au regard de ses objectifs, de caractériser la population cible, les parties impliquées dans sa mise en œuvre et sa temporalité. Cela implique par ailleurs de recenser les études internationales portant sur les évaluations de politiques similaires, mais réalisées dans d'autres contextes, afin de constituer une ressource pour le calibrage de l'analyse coûts-bénéfices. Il importe ensuite de caractériser précisément la situation contrefactuelle et les différentes options, et enfin de définir les problèmes liés à l'évaluation et les méthodes pour y répondre.

3.2. Décider des coûts et bénéfices à imputer aux différents types d'acteurs concernés par la politique

Pratiquement toutes les politiques publiques impliquent plusieurs types d'acteurs et de bénéficiaires ; chaque bénéfice, chaque effet, etc., affecte un groupe particulier de personnes, souvent les bénéficiaires directs de la politique envisagée. Mais c'est parfois l'ensemble des contribuables qui supporte le coût de la mise en œuvre de cette politique.

Puisque le calcul socioéconomique concerne la société dans son ensemble, l'analyse doit inclure l'ensemble des groupes de personnes ou d'institutions qui subissent les coûts ou perçoivent les bénéfices induits par la politique considérée. Cette définition doit être adaptée à la politique évaluée. Si celle-ci concerne un territoire extrêmement limité, il n'est pas nécessaire de considérer d'autres territoires, sauf si l'on s'attend à ce que des externalités, positives ou négatives, affectent certains groupes de personnes dans ces autres territoires.

3.3. Caractériser et quantifier les coûts et bénéfices

Après avoir identifié les coûts et les bénéfices de chaque groupe de parties prenantes qui doit être pris en compte dans l'analyse, l'étape suivante consiste à quantifier et caractériser ces coûts et bénéfices, autant que faire se peut. Dans cette étape, ce qui est défini comme coût et comme bénéfice n'a pas nécessairement une valeur monétaire marchande mais peut correspondre à un effet non directement monétaire de la politique évaluée. C'est pourquoi il est utile de recenser les coûts et bénéfices relevant ou non d'un marché, à savoir, d'un « mécanisme par lequel des acheteurs et des vendeurs interagissent pour déterminer le prix et la quantité d'un bien ou d'un service⁴⁴ ». Par ailleurs, les coûts et les bénéfices peuvent être caractérisés comme des effets directs ou

⁴⁴ P. Samuelson et W. Nordhaus (2000), *Economie*, 16^e édition, Paris : *Economica*, p. 27.

externes. Les effets directs sont ceux directement imputables à la politique, ou encore ceux entretenant un lien direct de cause à effet avec cette politique. Les effets externes sont ceux relevant d'ajustements à la politique et sont souvent appelés « effets de second tour ». Par convention, il est courant de parler d'effets négatifs d'une politique comme « coûts sociaux », et symétriquement des effets positifs comme des « bénéfiques », qu'ils interviennent lors de la mise en place de la politique ou qu'ils soient des conséquences de celle-ci. Cet abus de langage ne doit pas être mal interprété. Ce qui est considéré comme « coûts » dans une analyse coûts-bénéfices est **composé de la dépense différentielle pour la mise en œuvre de la politique**. Elle ne comprend pas les effets négatifs valorisés en unités monétaires. Pour plus de clarté, il est aussi possible de classer les coûts comme des intrants (*inputs*) et les bénéfiques comme des résultats (*outputs*).

Les effets sont ensuite quantifiés à partir d'une analyse reflétant les changements induits par la mise en œuvre de la politique par rapport à la situation contrefactuelle (qui est, en général, le *statu quo*). La quantification des effets suppose de disposer de méthodes et de travaux statistiques permettant de les identifier et de les estimer de la manière la plus précise possible.

Pour une évaluation *ex ante*, c'est la littérature existante qui fournira les estimations nécessaires (voir section suivante). Pour les évaluations *ex post*, ces données doivent être obtenues grâce au cadre de recherche élaboré au moment de la mise en œuvre.

Néanmoins, dans la plupart des cas, ces travaux n'existent pas, ou alors ils sont encore rares. Beaucoup de ces effets doivent alors être anticipés et/ou extrapolés à partir de données de moins bonne qualité et potentiellement biaisées, qu'il s'agisse de statistiques descriptives ou de dires d'experts. Par ailleurs, ces analyses sont fréquemment menées dans un cadre d'équilibre partiel, c'est-à-dire toutes choses égales par ailleurs. **Proposer un modèle plus complet, tenant compte des liens entre les différents marchés, revient à construire un modèle d'équilibre général, mais cette démarche est rarement possible ou utilisée dans l'évaluation socioéconomique.**

Tous les effets ne peuvent être estimés avec précision, soit parce que les relations de causalité ne sont pas observables (*i.e.*, l'information statistique est insuffisante), soit simplement parce qu'il est difficile d'exprimer ces effets en termes quantitatifs. **Les effets qui ne peuvent être quantifiés doivent néanmoins être décrits de façon qualitative.**

3.4. Projeter les coûts et bénéfiques sur toute la période

L'étape suivante consiste à analyser la façon dont les coûts et les bénéfiques vont se répartir et évoluer dans le temps. Les coûts et les bénéfiques ne sont généralement pas simultanés et s'échelonnent de manière différente au cours du temps. Les analyses coûts-bénéfices ou coûts-efficacité peuvent être conduites sur des horizons de durée très variable, allant de quelques mois, voire quelques années, à plusieurs décennies ou même à un cycle de vie entier. Le choix de l'horizon temporel est un choix aux conséquences souvent importantes sur l'estimation du rendement social de la politique considérée. Par conséquent, si plusieurs politiques sont comparées, il est important que

les calculs soient faits avec le même horizon temporel et que soit testée la sensibilité des résultats à la période considérée.

Dans le cas d'une analyse *ex ante*, il faut prévoir et anticiper l'ensemble des coûts et bénéfices imputables aux périodes futures. De nombreux choix méthodologiques doivent être alors faits : les coûts et les bénéfices se dissipent-ils ou bien sont-ils pérennes ? Ont-ils un effet multiplicateur ? Évoluent-ils de façon discontinue ou régulière ?

Pour les analyses *ex post*, on dispose en général de plus d'informations sur les bénéfices et les coûts mais rarement de séries complètes ou suffisantes permettant d'« observer et de reconstituer » directement leur évolution (cf., par exemple, l'analyse du programme de préscolarisation Perry Preschool présenté dans les annexes).

3.5. Estimer les coûts de l'investissement social

Cette étape consiste à évaluer les coûts de l'intervention, c'est-à-dire les coûts opérationnels définis par rapport à la situation contrefactuelle. Dans la mesure où la mise en œuvre de la politique est souvent gérée par d'autres institutions que celles en charge des décisions ou de l'évaluation, les coûts opérationnels sont rarement disponibles pour l'analyste qui doit alors les reconstruire. Des méthodes permettent de recenser et de reconstituer ces coûts, et ensuite de construire les indicateurs pertinents comme la méthode des ingrédients (voir Chapitre 3). **Cependant, ces coûts ne doivent concerner que les coûts différentiels liés au changement par rapport à la situation contrefactuelle.** Ainsi, les coûts d'utilisation du capital physique ne seront pas intégrés dans le calcul si le projet n'implique pas de modification de son utilisation. Il en est de même pour les salaires des personnels s'il est prévu qu'ils soient payés de la même façon dans la situation contrefactuelle. De ce fait, le coût opérationnel du programme est généralement très inférieur à celui que l'analyste financier ou le comptable a calculé.

Monétisation

L'étape suivante consiste à attribuer des valeurs monétaires aux effets quantifiés de la politique. Cette étape est celle qui est le plus souvent critiquée. En particulier parce que la notion même de valorisation monétaire (ou de monétisation) peut être remise en question sur la base de conceptions éthiques ou philosophiques : pourquoi donner une valeur monétaire au temps, à la vie, à la pauvreté, à la délinquance, au plaisir de la parentalité, etc. ?

La monétisation est vue ici seulement comme un moyen d'exprimer à l'aide d'une unité commune une réalité complexe à des fins de comparaison. Les valeurs utilisées pourraient apparaître contestables, voire non éthiques ; elles ne reflètent généralement pas une valeur affective, émotionnelle, ou substantielle. Toutefois, grâce à la monétisation de certains bénéfices, les analyses qui suivraient les recommandations de ce guide et utiliseraient les mêmes valorisations pourraient être comparées entre elles, facilitant ainsi les choix des décideurs.

Les bénéfices sont définis comme la valeur monétaire des effets causaux identifiés précédemment qui peuvent être de nature très variée. Pour les bénéfices, on trouve

souvent un niveau de diplôme plus élevé ou une meilleure réussite à un test, un taux d'emploi plus élevé, une espérance de vie plus élevée, etc. Dans les bénéfices négatifs (*i.e.* les « coûts sociaux évoqués précédemment), on trouve généralement une perte d'espérance de vie, une augmentation de la pollution, des comportements à risque plus fréquents, etc. Le coût social intègre également les **coûts d'opportunité** : si des ressources sont consacrées à une nouvelle politique, elles ne peuvent l'être à d'autres projets ou besoins publics ; des arbitrages doivent être faits. Les coûts d'opportunité incluent habituellement les coûts d'opportunité des fonds publics, les salaires non perçus des bénéficiaires en formation, etc. C'est là une caractéristique distinctive du raisonnement économique, qui se différencie ainsi de l'approche comptable.

Symétriquement, certains bénéfices prennent la forme de **coûts évités**. Par exemple, un modèle de microsimulation assez complet peut fournir un moyen d'estimer les bénéfices en termes d'assurance chômage, de minima sociaux, de transferts socio-fiscaux, mais aussi des bénéfices liés à des contributions fiscales et sociales potentiellement plus importantes si la politique évaluée permet d'accroître l'employabilité des bénéficiaires.

Pour tous ces effets, il faut pouvoir donner une valeur monétaire et cela n'est pas chose aisée, en raison notamment des désaccords sur certains de ces thèmes (voir chapitre suivant).

La monétisation repose sur des choix et hypothèses qui doivent être exposés, argumentés et testés par le biais d'analyses de sensibilité afin de déterminer à quel point les résultats de l'évaluation socioéconomique sont liés à ces hypothèses. Les valeurs sociales de ces coûts et bénéfices peuvent être obtenues de plusieurs façons. L'une des principales est l'évaluation contingente, méthode par laquelle des personnes sont amenées à exprimer leurs préférences et leur consentement à payer en vue d'obtenir un résultat. Ces valorisations peuvent être également construites statistiquement à l'aide d'une comptabilité analytique⁴⁵, ou bien encore inférées à partir de données statistiques et de modèles économiques permettant d'obtenir des préférences révélées⁴⁶. Elles peuvent enfin être estimées à l'aide de valeurs tutélaires ou conventionnelles sur la base de recensions de travaux existants ou de méta-analyses⁴⁷. À l'heure actuelle, il n'y a pas de méthode unique permettant de définir la valeur sociale d'un effet. Une section du chapitre 5 est consacrée à la discussion des méthodes de monétisation.

Mesure d'efficacité

Dans le cas d'une analyse coûts-efficacité, cette étape de valorisation monétaire n'est pas nécessaire : l'analyse consiste généralement à ne garder que l'unité (non monétaire)

⁴⁵ Par exemple, le coût évité d'un délit peut être estimé à partir des coûts d'utilisation du système judiciaire et pénitentiaire, de statistiques d'assurances caractérisant les dédommagements de certains délits, tels des vols ou les coûts de prise en charge de victimes d'agression ou de violence, etc. Il peut être complété par une évaluation contingente.

⁴⁶ La valeur d'un diplôme par exemple peut être obtenue par un modèle économétrique identifiant l'effet de ce diplôme sur le salaire.

⁴⁷ Voir les travaux de France Stratégie sur la vie humaine ou le CO₂. Rapport Quinet (2013), *L'évaluation socioéconomique des investissements publics*, tome 1.

de décompte du bénéfice principal comme unité d'efficacité. On raisonne alors sur une échelle naturelle (points gagnés au baccalauréat, etc.). Le secteur de la santé a pour sa part développé une mesure des bénéfices permettant d'aller plus loin que la seule mesure d'efficacité, et ce afin de comparer entre elles des interventions cliniques de nature différente. Il s'agit par exemple du QALY (*Quality adjusted life years*), développé par l'université de York dans les années 1980 et adopté par le NICE (*National Institute for Health and Clinical Excellence*) britannique, ou le DALY développé par l'OMS⁴⁸ pour estimer la charge mondiale de morbidité. Le QALY prend en compte les variations d'espérance de vie par une mesure du bien-être associé aux années futures, de sorte qu'une année en bon état de santé reçoit un poids de 1, la mort un poids de 0 et une année en mauvaise santé un poids entre 0 et 1. Les poids sont estimés à partir de questionnaires et différenciés par pays. Certains poids peuvent même être négatifs dans le cas où le patient malade préférerait la mort à la survie dans un état végétatif ou en grande souffrance⁴⁹. Le recours à ce type d'indicateur plutôt qu'à une ACB tient principalement au refus de donner une valeur monétaire à des résultats de santé pour des raisons éthiques. La Haute Autorité de Santé (HAS) a souligné dans un rapport de 2014 le besoin qu'elle avait de pouvoir utiliser des analyses coûts-utilité (basées sur les Qalys) pour mener à bien la mission d'évaluation médico-économique qui lui a été confiée en 2012 par la loi de financement de la Sécurité sociale, en produisant des ratios différentiels coûts-résultats qui comparent la nouvelle stratégie de santé évaluée à la pratique courante⁵⁰.

3.6. Actualiser les coûts et les bénéfices

Les bénéfices et les coûts d'une politique d'investissement social interviennent à des moments différents. Afin de pouvoir réaliser des calculs monétaires à partir de flux intervenant à des dates différentes, il est nécessaire d'exprimer ces flux à une date commune, autrement dit de les « actualiser ». Le taux d'actualisation permet de ramener à une même date des flux monétaires intervenant à des années différentes, et donc de faire l'arbitrage entre le présent et le futur. L'actualisation reflète l'idée que même sans inflation, 100 euros aujourd'hui ont plus de valeur que 100 euros dans un an, et beaucoup plus que 100 euros dans dix ans. Le taux d'actualisation illustre donc la « préférence pour le présent⁵¹ ». Cette règle s'applique également pour les coûts et bénéfices d'un investissement social.

Pour tenir compte de ces effets de valorisation du temps, on convertit les valeurs à chaque date en les incrémentant ou décrémentant d'un facteur d'actualisation, résultant de la prise en compte du taux d'actualisation sur un plus ou moins grand nombre de périodes, selon la distance à la date de référence du calcul.

⁴⁸ Le DALY est présenté plus précisément dans l'annexe A-III.

⁴⁹ Chamot E., Bovier P., Perneger T., Chastonay P., Herrmann F., Garnerin P. et Chopard P. (2001), « Analyse économique appliquée à la santé : un tour d'horizon », *Bulletin des médecins suisses*, Vol. 82(39).

⁵⁰ Haute Autorité de santé (2014), « Valeurs de références pour l'évaluation économique en santé, Revue de la littérature », Document de travail, HAS, décembre.

⁵¹ Plus ce taux est élevé, moins l'individu accorde de valeur aux bénéfices et aux coûts éloignés dans le temps.

3.7. Calculer le ratio coûts-efficacité ou coûts-bénéfices

Cette étape constitue le cœur du calcul socioéconomique. Elle consiste à agréger les valeurs actualisées des coûts et des bénéfices, ou celles des coûts et des efficacités.

Le ratio coûts-efficacité est simplement le ratio entre la mesure d'efficacité considérée et la somme des coûts actualisés. Il s'exprime en euros par unité gagnée. Son inverse peut parfois être utilisé ; il s'interprète comme l'effet d'un euro (ou d'un millier d'euros) supplémentaire dépensé. Les comparaisons entre des ratios coûts-efficacité ou efficacité-coûts doivent être faites avec une certaine prudence. Par exemple, un programme peut coûter dix fois plus qu'un autre programme dix fois moins efficace, leurs ratios coûts-efficacité seront voisins, alors que leurs coûts et bénéfices sont en réalité très différents. Les ratios coûts-efficacité sont donc plutôt utiles pour comparer des projets de dimensions semblables.

Dans le cas des analyses coûts-bénéfices, la principale mesure est la **valeur actualisée nette socioéconomique (VAN-SE)** qui est la somme actualisée des coûts et des bénéfices sociaux. Deux autres statistiques sont souvent produites :

- le ratio bénéfices-coûts ou bénéfice par euro dépensé, qui est le ratio entre la somme actualisée des bénéfices et la somme actualisée des coûts. Ce ratio est intéressant car il facilite à la fois la comparaison de programmes différents et la compréhension des résultats du calcul puisque tout ratio supérieur à 1 correspond à une politique efficiente dans laquelle les bénéfices l'emportent sur les coûts. Cependant, comme pour les ratios coûts-efficacité, la comparaison doit se faire pour des projets de dimensions comparables car les effets d'échelle sont ici également présents ;
- le taux de rendement interne (TRI), qui est le taux d'actualisation qui annule la VAN-SE.

3.8. Produire une analyse de sensibilité

Le calcul socioéconomique repose sur de nombreuses hypothèses mais aussi sur l'utilisation d'estimations ou de calibrations issues de travaux eux-mêmes fondés sur des hypothèses. En conséquence, les méthodes d'estimation des rendements socioéconomiques augmentent significativement le risque de présenter des résultats aberrants parce qu'elles cumulent, parfois entre elles, des valeurs incertaines (certains statisticiens parleraient seulement d'erreurs de mesure). Par exemple, un calcul utilisant des valeurs moyennes peut indiquer qu'un programme d'investissement social rapporte à la société 2,50 euros par euro dépensé. La question fondamentale est de savoir à quel point cette estimation est risquée. Si l'on modifiait les valeurs des paramètres, est-ce que les bénéfices resteraient supérieurs aux coûts, ou l'inverse ?

Toutes les évaluations socioéconomiques sont basées sur des hypothèses qu'il faut tester, mais la spécificité de l'investissement social tient à un recours accru à des paramètres estimés économétriquement.

Il est alors crucial de tester la sensibilité des résultats aux hypothèses faites sur les valeurs utilisées pour simuler ou calibrer le modèle de calcul. **Deux types d'analyses de**

sensibilité sont communément pratiquées : 1) la méthode des cas extrêmes – qui consiste à n'utiliser que les estimations les plus défavorables (ou les plus favorables) et voir si le résultat final reste raisonnable ; et 2) les analyses qui consistent à imputer d'autres valeurs que celles initialement retenues.

Mais une troisième méthode est parfois préférable : il s'agit d'une analyse de type Monte-Carlo qui consiste à répliquer plusieurs centaines, voire plusieurs milliers de fois le calcul coûts-bénéfices en tirant au hasard les paramètres du modèle dans des distributions de valeurs plausibles. Cette méthode n'est pas sans défauts car elle repose sur des hypothèses sur la forme et les paramètres (moyenne, variance, etc.) de ces distributions. Elle a toutefois l'intérêt de fournir, en plus du rendement social moyen, une distribution et une mesure de l'incertitude de son estimation. En comptant par exemple le nombre de fois où la VAN-SE est positive parmi le nombre total de réplifications, on obtient une estimation de la probabilité que le rendement de la politique évaluée soit positif.

3.9. Présenter les résultats et faire des recommandations

La dernière étape de l'évaluation socioéconomique consiste à présenter les résultats et éventuellement à faire des recommandations.

Pour que les résultats de l'analyse socioéconomique puissent aider à la décision et à la mise en œuvre d'une politique d'investissement social, ils doivent être **présentés de façon claire**. Dans l'idéal, ils devraient être reproductibles. Les principaux résultats peuvent être présentés dans un tableau et sous forme de graphiques. Les principales variantes doivent également être reportées : elles doivent présenter les VAN-SE correspondantes, ainsi qu'une **mesure de l'incertitude pour chaque VAN-SE** telle que le ratio du nombre de fois où la valeur de la VAN-SE est positive sur le nombre total de simulations. **La monétisation de certains effets, les incertitudes inhérentes au calcul socioéconomique et les principales hypothèses faites pour pallier ces problèmes doivent être discutées.**

Le raisonnement qui fonde la construction du modèle est tout aussi important que les résultats eux-mêmes. Les décideurs et les citoyens doivent connaître les choix faits dans le cadre de la formalisation, mais aussi les limites de l'exercice afin de pouvoir conserver un recul critique à son égard, et ce faisant une faculté de jugement et de choix.

Les analystes chargés de l'évaluation socioéconomique peuvent proposer des recommandations à partir des estimations obtenues parce qu'elles ou ils disposent d'une information privilégiée et complète sur la plausibilité des hypothèses faites, la robustesse de leurs calculs et l'état de la connaissance sur l'efficacité des politiques évaluées. Toutefois, leurs recommandations doivent être transparentes, ce qui garantira notamment leur caractère non partisan (objectivation des hypothèses) et leur humilité (objectivation des incertitudes).

Que retenir ?

Conduire une analyse coûts-bénéfices consiste en une série d'étapes successives. Les guides existants ne préconisent pas de réaliser toutes ces étapes dans le même ordre. En particulier, on se doit de mettre l'accent sur la distribution des effets entre parties prenantes, sur l'existence d'effets hétérogènes et sur leur répartition dans le temps. C'est pourquoi nous nous éloignons sensiblement des étapes préconisées dans le rapport de France Stratégie (2017).

Pour l'investissement social, les dix étapes d'une analyse coûts-bénéfices sont :

1. *Définir le cadre général d'analyse*
2. *Décider des coûts et bénéfices à imputer aux différents types d'acteurs concernés par l'analyse*
3. *Identifier et caractériser ces coûts et bénéfices*
4. *Prévoir l'évolution des coûts et des bénéfices tout au long de la mise en œuvre de la politique*
5. *Évaluer les coûts de l'investissement social*
6. *Quantifier les bénéfices par le biais d'une mesure de l'efficacité (pour une analyse coûts-efficacité) ou monétiser les bénéfices (pour une analyse coûts-bénéfices)*
7. *Actualiser les coûts et bénéfices futurs au moyen d'une valeur actualisée*
8. *Calculer un ratio coûts-efficacité (pour une analyse coûts-efficacité) ou une valeur actualisée nette socioéconomique (pour une analyse coûts-bénéfices)*
9. *Conduire une analyse de sensibilité*
10. *Faire des recommandations appropriées*

4. Quelles études sélectionner pour une revue de littérature ?

4.1. La validité des évaluations antérieures

L'étude d'impact est une étape antérieure à l'élaboration d'une ACB. Une évaluation d'impact doit avoir au moins deux qualités. La première est son degré de « validité externe ». Les évaluations de politiques similaires sont souvent réalisées dans des pays, des périodes ou des contextes différents. La transposabilité de leurs résultats n'est donc pas certaine. La seconde qualité exigée d'une évaluation est sa « validité interne ». La méthodologie d'évaluation est-elle suffisamment rigoureuse et permet-elle d'identifier des relations causales ? Est-elle exempte de biais de sélection ? L'échantillon est-il suffisamment grand et représentatif ? Est-on sûr du sens de la causalité ? N'existe-t-il pas d'autres facteurs qui ne peuvent être observés et qui soient cause des phénomènes observés ? Les progrès accomplis ces dernières années dans le champ des méthodes ont mis l'accent sur les enjeux de validité interne et ont conduit à déclasser un grand nombre d'études passées. Pour réaliser une analyse coûts-bénéfices, il est indispensable de définir au préalable les critères de qualité scientifique qui permettent de retenir certaines études et d'en ignorer d'autres.

4.2. Éviter les revues partielles de littérature

On pourrait être tenté de choisir des études dont la méthodologie semble robuste et qui ont été réalisées dans des contextes semblables à celui dans lequel va être mise en place la politique considérée. C'est généralement ce qui est fait et ce qui semble le plus raisonnable. Cette démarche conduit toutefois le chercheur à faire des choix dont les critères sont flous. Par exemple, est-on certain de la similitude des contextes ? Ce problème de « *cherry picking* » (littéralement, de « picorage » d'indices anecdotiques) est renforcé lorsque les critères sont définis au fur et à mesure de la recherche bibliographique ou lorsque celle-ci n'est pas subordonnée à une règle d'arrêt.

En outre, des études récentes ont montré qu'une grande partie des résultats d'évaluations d'impact qui sont publiées dans des revues universitaires ne peuvent être répliqués (et ce parfois avec les mêmes jeux de données). Ce défaut a plusieurs causes⁵² :

- tout d'abord, il existe un biais de publication : les études dont les résultats sont statistiquement non significatifs (c'est-à-dire, celles qui ne peuvent conclure à l'existence d'effets), ne sont généralement pas aussi facilement acceptées pour publication dans les revues avec comité de lecture que celles qui trouvent des résultats significatifs ;
- en conséquence, les résultats qui concluent à l'absence d'effets restent malheureusement souvent « dans le tiroir » et personne ne sait qu'ils existent⁵³ ;
- les chercheurs sont fortement incités à adapter leurs méthodes afin d'obtenir des résultats statistiquement significatifs (ce phénomène est appelé le *p-hacking*)⁵⁴ ;
- les modèles sont généralement élaborés une fois le résultat observé, de sorte que ces modèles correspondent *a posteriori* aux données (c'est le phénomène de *harking*).

L'article de Brodeur *et al.* (2018)⁵⁵ recense les statistiques de test reproduites dans les articles publiés dans les 25 meilleures revues d'économie en 2015 en fonction de la méthode d'évaluation utilisée⁵⁶. La Figure 3 en est extraite : elle montre qu'à l'exception des techniques de régression avec discontinuité, les méthodes d'évaluation d'impact les plus rigoureuses présentent plus fréquemment des statistiques de test voisines du seuil habituel de significativité, ce qui peut être interprété comme une forme de *p-hacking*.

⁵² Une discussion de ces enjeux est proposée de façon très convaincante par Sylvain Chabé Ferret [sur son site personnel](#). Les éléments présentés dans cette section s'en inspirent grandement.

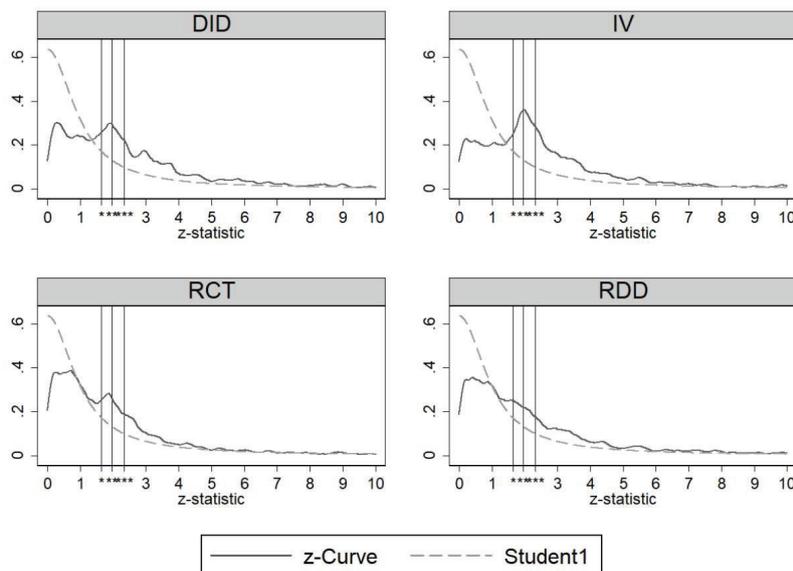
⁵³ Ce point concerne à la fois les études qui ne parviennent pas à être publiées, mais aussi et surtout les analyses peu concluantes qui ne sont même pas référencées dans les documents de travail. Une façon de limiter ce problème consiste à déclarer les protocoles et les plans de préanalyse (voir *infra*). Il est également possible de rendre public un « *log file* », c'est-à-dire un fichier reprenant toutes les étapes de l'analyse.

⁵⁴ Comme pour le point précédent, un plan de préanalyse permet d'éviter ce problème.

⁵⁵ Brodeur A., Cook N. et Heyes A. (2018), "Methods matter: P hacking and causal inference in economics", IZA Discussion paper, No. 11796, Bonn.

⁵⁶ DiD correspond à la méthode des doubles différences, IV correspond à l'utilisation d'une ou plusieurs variables instrumentales, RCT correspond à une expérience aléatoire contrôlée et RDD correspond à une discontinuité de régression.

Figure 3 – Densité des statistiques de test issus d'articles publiés dans les 25 meilleures revues d'économie en 2015



Source : d'après Brodeur et al. (2018)

Ainsi, une revue de littérature sélective, puisant ses exemples dans de telles publications, peut avoir des conséquences malencontreuses, notamment lorsqu'elle est utilisée pour guider la conception d'une nouvelle politique publique, car elle laisse croire qu'une politique semblable à celle envisagée, lorsqu'elle a été mise en œuvre par le passé dans des contextes similaires, a eu des effets positifs. La section 2.IV.4 revient sur cette question.

La façon la plus rigoureuse de sélectionner les études à utiliser est donc de procéder à une revue systématique de littérature basée sur des critères d'inclusion explicites, définis au préalable, et adaptés à la politique d'investissement social considérée.

Les critères de sélection des études doivent concerner :

- la **qualité du protocole d'évaluation** : retenir les expérimentations randomisées ou les études réalisées à partir de méthodes quasi-expérimentales tenant compte des caractéristiques non observables ;
- les **sources** : privilégier non seulement les revues universitaires avec comité de lecture, mais aussi les documents de travail publiés par des centres de recherche, la littérature grise, etc. ;
- des **périodes récentes et des pays pertinents pour la comparaison** : retenir principalement les études relatives à des pays européens, de l'OCDE, etc., publiées durant les dix à trente dernières années ;
- les **mots clés** mis en avant dans les recherches bibliographiques : définir les mots clés prioritaires, dater le moment où la littérature est collectée.

Une revue de littérature systématique doit être exhaustive, ce qui implique un travail conséquent de recension, d'analyse et de tri pour supprimer les doublons et ne conserver

que les travaux pertinents. C'est un exercice coûteux, notamment parce la plupart des revues académiques sont payantes et leur accès est restreint⁵⁷. Afin de réduire les coûts, il est possible de réaliser un premier tri à partir des résumés d'articles pour ne garder que ceux qui sont les plus pertinents. Il est possible de se servir pour cela des revues de littérature réalisées souvent dans des pays différents ou par des équipes qui produisent des revues de littérature identiques⁵⁸. Il faut noter que les revues de littérature effectuées dans les milieux médicaux ont des exigences très élevées et sont très en avance en termes de méthodes par rapport à celles réalisées par les économistes. C'est pourquoi les revues systématiques et méta-analyses sont beaucoup plus développées en santé qu'ailleurs, même si l'émergence d'institutions comme Campbell Collaboration démocratise leur usage en sciences sociales (voir Annexe B).

4.3. Développer les méta-analyses

La recension exhaustive de littérature peut dans certains cas permettre de réaliser une méta-analyse, c'est-à-dire une analyse statistique des résultats collectés tenant compte de la précision statistique des résultats, du nombre d'observations et des biais potentiels. Dans une méta-analyse, les résultats des études retenues sont standardisés⁵⁹, pondérés en fonction du nombre d'observations et d'autres critères (par exemple, l'évaluation du risque de biais statistique), de façon à obtenir un effet moyen méta-analytique. Il est souvent nécessaire de faire des ajustements statistiques car les modèles utilisés et les valeurs ne sont pas identiques. Le manuel de Wipsey et Wilson (2001)⁶⁰ fournit des explications très claires sur la démarche à suivre et les méthodes à employer suivant les situations⁶¹.

Une méta-analyse a pour principal intérêt de synthétiser la connaissance disponible à un instant donné, indépendamment des lieux et des périodes où les politiques ont été mises en place⁶². Lorsqu'elle repose sur un nombre important d'études, l'effet synthétisé qu'elle délivre peut être considéré comme l'effet moyen hors effets de contexte. Sa validité externe est vraisemblablement plus grande. Par ailleurs, les méta-analyses permettent de détecter les biais de publication. En présence d'un tel biais, on s'attend à ce qu'il y ait plus de conclusions positives que négatives et des effets de plus en plus élevés à mesure que la précision décroît.

⁵⁷ Cela n'est généralement pas le cas pour les documents de travail non publiés et la littérature grise.

⁵⁸ Voir les revues de littérature de Cochrane (<https://www.cochranelibrary.com/>) et de Campbell (<https://campbellcollaboration.org/>).

⁵⁹ Une variable de résultat est standardisée lorsqu'elle est rapportée à son écart-type, de façon à obtenir des tailles d'effet comparables pour des variables différentes. Cette variable standardisée est appelée le D de Cohen.

⁶⁰ Lipsey M. W. et Wilson D. (2001), *Practical Meta-Analysis*, Sage Publisher.

⁶¹ Des outils pratiques sont également disponibles sur le site <http://mason.gmu.edu/~dwilsonb/ma.html>, avec notamment un convertisseur de taille d'effet. Une version en ligne est disponible sur le site de Campbell collaboration : <https://www.campbellcollaboration.org/research-resources/research-for-resources/effect-size-calculator.html>.

⁶² Gurevitch J., Koricheva J., Nakagawa S. et Stewart G. (2018), "Meta-analysis and the science of research synthesis", *Nature*, Vol. 555, mars, p. 175-182.

Encadré – Un exemple de méta-analyse : l'effet de la taille des classes sur la réussite scolaire

Fildge *et al.* (2018)⁶³ ont réalisé une méta-analyse des études consacrées à l'effet de la taille des classes sur les résultats scolaires des élèves à l'école primaire. En utilisant les bases de données bibliographiques en ligne disponibles en février 2017, les auteurs ont recensé les travaux publiés dans des revues à comité de lecture, dans les séries de documents de travail, dans la littérature grise, etc., sans restriction quant au pays où avait eu lieu l'évaluation. Les études devaient être quantitatives, concerner la réduction de la taille des classes, produire des mesures de réussite scolaire et utiliser une méthodologie de qualité. Ils ont ainsi identifié ainsi 8 128 études parmi lesquelles 127 respectaient les critères d'inclusion. Ces 127 études portent sur 55 échantillons provenant de 41 pays différents. Parmi ces 127 études, 45 concernent la seule expérience *STAR* (voir le Chapitre 4 pour une présentation détaillée de cette expérimentation).

Parmi les 82 études restantes, 58 n'ont pas été intégrées dans la synthèse parce qu'elles étaient sujettes à un biais de sélection potentiellement important ; 18 autres études ne permettaient pas de calculer une taille d'effet. Au final, seules 6 études (dont 2 françaises) en dehors de celles de *STAR* ont pu être incorporées dans cette méta-analyse.

Figure 4.1: Reading

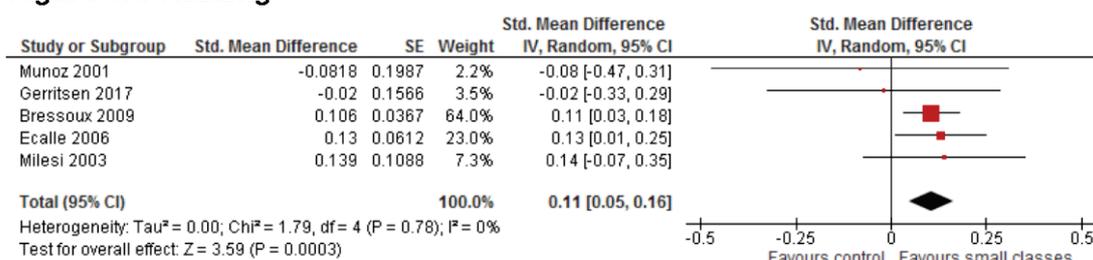
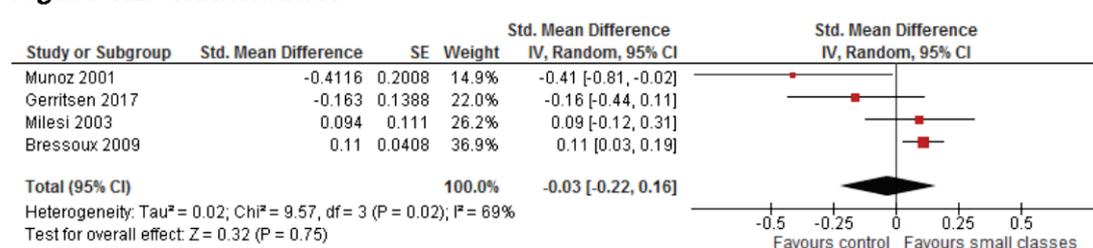


Figure 4.2: Mathematics



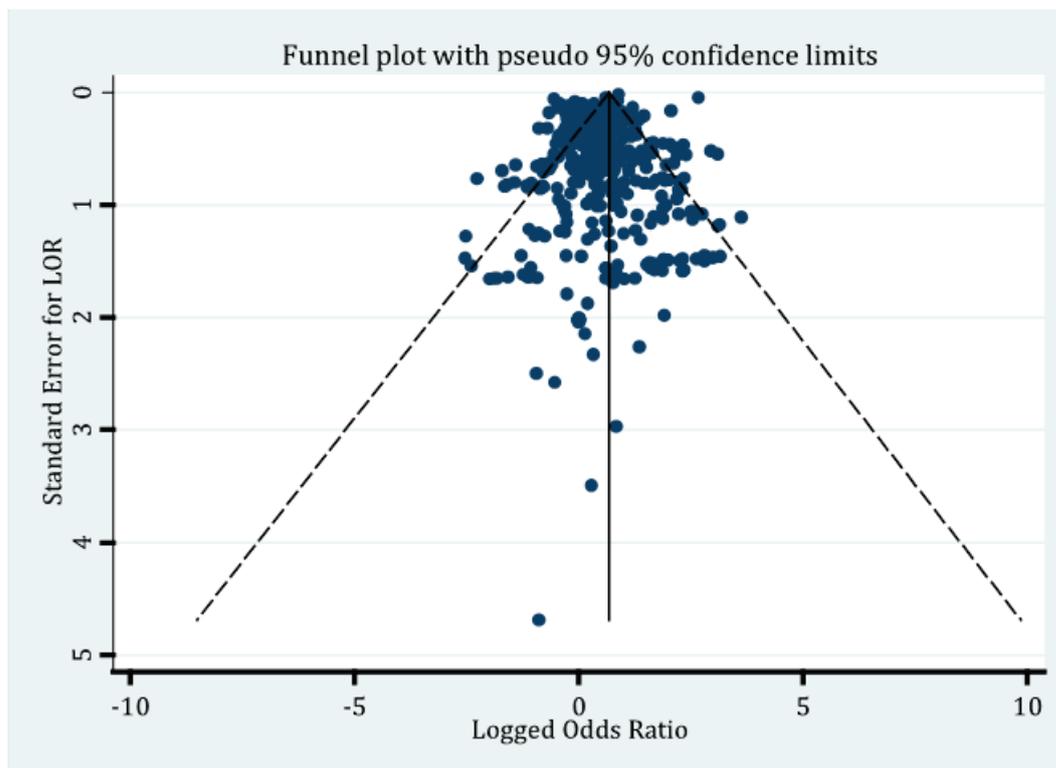
Les résultats indiquent un effet positif (compris entre 0,05 et 0,16) de la réduction de la taille de la classe sur les résultats en lecture, mais pas d'effet en mathématiques. Les résultats sur l'expérience *STAR* permettent de conclure à un effet positif significatif à la fois en lecture et en mathématiques.

Les graphiques en « entonnoir » (*funnel*) sont un outil particulièrement utile : ils représentent sur l'axe des abscisses les tailles d'effet et en ordonnées les erreurs standards associées aux coefficients. Un triangle est ensuite tracé autour de l'effet moyen résultant de la méta-analyse, les côtés de ce triangle correspondant aux limites de la significativité statistique. En

⁶³ Filges T., Sonne-Schmidt C. S. et Nielsen B. C. V. (2018), "Small class sizes for improving student achievement in primary and secondary schools", *Campbell Systematic Reviews*, 2018:10.

l'absence de biais de publication, les effets recensés doivent être distribués symétriquement autour de la taille d'effet moyenne et être situés autant à l'intérieur qu'à l'extérieur du triangle, comme l'illustre la Figure 4 extraite de la méta-analyse sur les programmes de lutte contre le décrochage scolaire rédigée par Wilson *et al.* (2011)⁶⁴. Si les études sont concentrées à l'extérieur de l'entonnoir, c'est que seuls les résultats significatifs ont été publiés. S'ils ne sont pas répartis symétriquement, c'est que seuls ceux positifs (ou négatifs) ont été présentés.

Figure 4 – « Entonnoir » extrait de la méta-analyse sur les effets de prévention du décrochage réalisée par Wilson *et al.* (2011)



Avant de débiter une analyse coûts-bénéfices, il est donc recommandé de savoir s'il existe des méta-analyses qui synthétisent les effets estimés de politiques similaires à celle que l'on souhaite mettre en place. Certaines institutions internationales, comme la Campbell Collaboration ou Cochrane dans le domaine de la santé, ont contribué à financer, produire et diffuser de nombreuses méta-analyses. Le Washington State Institute for Public Policy (WSIPP) a également produit des méta-analyses relatives à de nombreuses politiques sociales. Toutefois, les méta-analyses du WSIPP concernent exclusivement des études nord-américaines et les critères d'inclusion y sont assez peu stricts.

⁶⁴ Wilson S. J., Tanner-Smith E., Lipsey M., Steinka-Fry K. et Morrison J. (2011), "Dropout prevention and intervention programs: Effects on school completion and dropout among school- aged children and youth", *Campbell Systematic Reviews*, 2011:8.

4.4. Améliorer la qualité des évaluations d'impact

Les évaluations des investissements sociaux promues tout du long de ce rapport peuvent être sujettes aux mêmes défauts que celles existantes. C'est pourquoi il importe de préconiser les pratiques de recherche les plus intègres.

En premier lieu, beaucoup de travaux de recherche présentent peu, ou mal, le contenu et les modalités de la politique qui est évaluée. Il est alors très difficile de comprendre la réalité de l'intervention et de la comparer à d'autres menées par ailleurs. Il est recommandé de rendre disponibles des éléments décrivant l'intervention, notamment sous forme d'annexes.

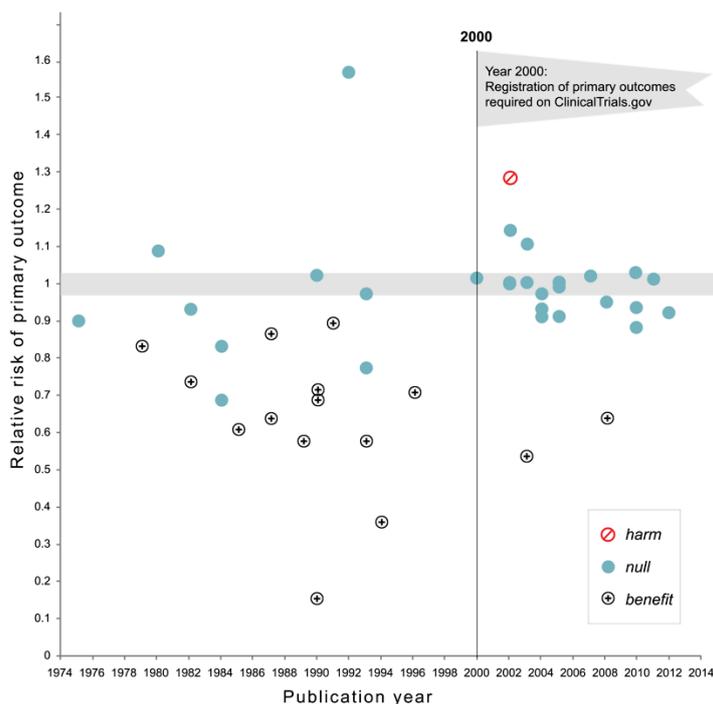
Ensuite, les procédures d'analyse (les codes) et les données utilisées, une fois purgées de tout élément identifiant les individus échantillonnés, devraient être accessibles, de sorte que les résultats puissent être répliqués à l'identique par des chercheurs tiers, et réanalysés le cas échéant. Même si toutes les données ne peuvent être rendues accessibles, notamment parce que certaines sont confidentielles ou parce qu'elles sont la propriété de leurs producteurs, la règle générale devrait être de les rendre publiques et de ne souffrir que peu d'exceptions.

Par ailleurs, les sections précédentes ont insisté sur l'importance de l'articulation *ex ante* et *ex post*, et sur la définition des conditions de collecte d'information et d'évaluation. Si cette étape est bien menée, les besoins de données et les méthodes d'analyse sont *de facto* connus. Dans ces conditions, il devrait être possible de rédiger un programme de préanalyse⁶⁵ (*pre-analysis plan*) décrivant précisément les étapes et le protocole de recherche. Il s'agit ici de spécifier les variables utilisées, les procédures pour vérifier et recoder les données, les modèles statistiques utilisés, etc. Si les modèles statistiques sont spécifiés à l'avance, il devient alors plus difficile pour l'analyste d'adopter des comportements douteux, comme le *harking* ou le *P-hacking*. Il ne s'agit pas d'une pratique nouvelle en science ; les évaluations d'impact dites « confirmatoires » – menées en épidémiologie afin d'obtenir la mise sur le marché de médicaments ou de produits alimentaires – doivent obligatoirement être dotées d'un programme de préanalyse. Aux États-Unis, l'effet de cette obligation sur les résultats publiés a été très fort, comme l'atteste la Figure 5.

Avant l'année 2000, lorsque l'enregistrement des plans de préanalyse n'était pas obligatoire, la plupart des résultats d'expériences étaient positifs et significatifs. Mais dès que les plans de préanalyse ont été rendus obligatoires, la plupart des résultats se sont révélés nuls. Il n'y a pas de raison de penser que depuis 2000 les nouvelles molécules testées sont systématiquement moins efficaces : c'est donc que beaucoup des résultats précédents reflétaient des pratiques de recherche douteuses, que l'enregistrement d'un plan de préanalyse a permis de corriger.

⁶⁵ Olken B. (2015), "Promises and perils of Pre-analysis plans", *Journal of Economic Perspectives*, V. 29, n°3, p 61-80.

Figure 5 – Résultats d'expériences aléatoires contrôlées financées par le National Heart Lung and Blood Institute (NHLBI) aux États-Unis entre 1974 et 2012



Source : Kaplan R.M. et Irvin V.L. (2015), "Likelihood of null effects of large nhlbi clinical trials has increased over time", *PLoS ONE*, Vol. 10(8): E0132382.

De plus, le programme de préanalyse permet de faciliter la publication des résultats. En effet, certaines revues⁶⁶ proposent d'évaluer la qualité de l'analyse sur la base du *pre-analysis plan* et, si celui-ci convient, de publier les résultats même si ceux-ci sont décevants.

A minima, si une expérimentation est mise en place, celle-ci doit être déclarée dans un registre, tel que celui proposé par l'American Economic Association⁶⁷ afin de limiter le problème du biais de publication.

Pour autant, un programme de préanalyse n'est pas indispensable. Il peut paraître assez inadapté dans certains cas. À ce sujet, la principale critique émise par Olken (2015) concerne la difficulté à anticiper toutes les possibilités. Il est en effet probable que pour des investissements sociaux affectant un nombre important de résultats, un plan de préanalyse complet soit hors de portée. Néanmoins, lorsque l'évaluation *ex post* est prévue en amont et conçue entièrement en vue de son analyse coûts-bénéfices, Le programme de préanalyse est déjà partiellement établi. L'effort nécessaire pour le finaliser est alors bien moindre.

En outre, l'évaluation d'investissements sociaux se fait généralement en lien étroit avec de nombreux partenaires, ce qui peut avoir pour effet de créer des situations délicates lorsque les résultats ne sont pas ceux espérés. Le plan de préanalyse peut alors permettre aux

⁶⁶ *Journal of development economics*, *Journal of experimental political sciences*, et de nombreux journaux en psychologie. Voir <https://cos.io/rr/>

⁶⁷ <https://www.socialscienceregistry.org/>

chercheurs et aux autres parties prenantes de se mettre d'accord sur la démarche avant que les résultats ne soient observés. Il a également pour vertu de protéger les chercheurs de possibles pressions ou incitations à remettre des résultats plus favorables.

Que retenir ?

La qualité d'une recherche est fonction de la rigueur de sa méthodologie (validité interne) et des possibilités d'extrapoler ses résultats à d'autres contextes ou populations (validité externe).

Les revues de littérature s'appuyant uniquement sur les travaux publiés risquent d'être biaisées car de nombreuses études ayant produit des résultats non significatifs ne sont pas publiées. À cela s'ajoutent les problèmes de reproductibilité et des pratiques de recherche douteuses.

Pour limiter le risque de sélection d'études, il est recommandé de mettre en place des revues systématiques de littérature à partir de critères clairs et préalablement définis.

À partir de ces revues systématiques, il est possible de réaliser des méta-analyses pour synthétiser les connaissances et détecter d'éventuels biais de publication.

Des institutions se sont spécialisées dans la production de travaux de ce type. Les méta-analyses déjà publiées peuvent servir à alimenter les calculs socioéconomiques ex ante.

Les évaluations ex post doivent être de qualité. Pour cela il convient de bien décrire l'investissement social considéré, de déclarer l'évaluation qui va être menée dans un registre public et, si possible, de fournir un plan de préanalyse.

Chapitre 3

Le cadre général de l'application du calcul socioéconomique aux politiques d'investissement social

1. La valeur actualisée nette

L'élément central, la pierre angulaire, du calcul socioéconomique, qu'il concerne l'investissement social ou non, est la valeur actualisée nette (VAN ci-après). Ce type de politique concernant généralement des personnes (des travailleurs, des chômeurs, des élèves ou des patients), l'échelle première de calcul est le niveau individuel. Une fois cette VAN calculée pour une personne en particulier, par exemple compte-tenu de son genre, de son âge, de son niveau d'éducation, de son origine sociale, etc., il conviendra d'agréger les valeurs individuelles pour établir une valeur actualisée nette globale qui tiendra compte de la structure de la population des bénéficiaires (en l'occurrence, proportions de femmes et d'hommes, de jeunes et de moins jeunes, etc., dans cette population). À ce stade, il conviendra également de tenir compte des coûts fixes globaux qui accompagnent la mise en œuvre générale de cette politique. Nous reviendrons sur cette question dans la section suivante.

L'élément principal d'une analyse coûts-bénéfices pour un individu, un salarié, un élève, etc., est la différence de deux valeurs actualisées :

$$VAN = V(1) - V(0)$$

$V(1)$ étant la valeur actualisée lorsque l'investissement social est réalisé et $V(0)$ celle associée à la situation contrefactuelle, c'est-à-dire ce que serait le monde sans l'investissement social. Ces deux valeurs sont de nature très théorique et ne sont pas identifiées. Seule la différence peut l'être dans les conditions décrites dans ce guide. L'intérêt de les définir est d'introduire un raisonnement contrefactuel et de statuer explicitement sur les hypothèses faites pour pouvoir calculer la valeur actualisée nette. En particulier, selon que l'on prend ou non en compte certains bénéfices, la situation sans l'investissement social est, ou non, modifiée. La VAN qui est obtenue par ce raisonnement différentiel peut être identifiée en se limitant aux variables pertinentes. Elle aura pourtant une validité générale sous l'hypothèse qu'il n'y pas eu d'externalités ou d'effets de second tour.

Cette définition impose alors de prendre en compte tous les changements induits par l'investissement social considéré.

Que retenir ?

L'élément central du calcul socioéconomique appliqué à un investissement social est la valeur actualisée nette (VAN).

Elle est définie comme la différence de deux valeurs actualisées, celle correspondant à la situation après l'investissement social et celle associée à la situation contrefactuelle, c'est-à-dire celle qui prévaut en l'absence de cet investissement social.

La VAN est obtenue par ce raisonnement différentiel. Elle est identifiée en se limitant aux variables pertinentes.

2. Les principaux paramètres

2.1. Le taux d'actualisation

Le taux d'actualisation joue ici le même rôle que dans le cas des investissements publics. Le *Guide de l'évaluation socioéconomique des investissements publics*⁶⁸ précise que « le taux d'actualisation permet de ramener à une même date des flux monétaires intervenant à des années différentes, et donc de faire l'arbitrage entre le présent et le futur ». En d'autres termes, le taux d'actualisation est utilisé pour déterminer la valeur actuelle des flux futurs (de revenus ou de coûts), c'est-à-dire leur valeur à une date donnée. Il s'interprète donc comme un taux mesurant la « préférence pour le présent » ; il est donc positif et permet de calculer la valeur à la période présente d'un flux financier de montant y_1 intervenant, par exemple, à la période suivante. Si l'on note r le taux d'actualisation, la valeur actualisée de ce revenu à la période présente est définie par :

$$VA_1 = \frac{y_1}{(1+r)}$$

Si le revenu perçu au cours de la t -ième période suivant la période présente est égal à y_t , sa valeur actualisée (*i.e.*, à la période présente) est égale à :

$$VA_t = \frac{y_t}{(1+r)^t}$$

La Figure 6 représente VA_t pour $y_t = 1000$, différentes valeurs de r et pour t allant de 0 à 50 ans. Ainsi, au taux de 4,5 %, 1000 euros à recevoir dans dix ans valent aujourd'hui environ 650 euros ; 750 si le taux est de 3 % et 550 euros si le taux est de 6 %. Un taux d'actualisation de 1 % fait que 1000 euros dans dix ans valent 950 euros aujourd'hui environ. Plus r est élevé, moins on accorde de valeur aujourd'hui à des revenus éloignés dans le futur.

En conséquence, lorsque la politique sociale ou éducative que l'on souhaite évaluer est susceptible d'avoir des effets sur les revenus futurs d'un individu jusqu'à la période T de son

⁶⁸ Direction générale du Trésor et France Stratégie (2017), *Guide de l'évaluation socioéconomique des investissements publics*, p. 34.

existence,⁶⁹ la valeur de ses revenus futurs actualisée au moment présent, qui appartient en général à la période de mise en œuvre de cette politique, est égale à la somme de ces revenus actualisés au taux d'actualisation :

$$VA_{[1,T]} = \sum_{t=1}^T \frac{y_t}{(1+r)^t}$$

Ainsi, les bénéfices futurs d'une politique sociale seront d'autant moins valorisés que le taux d'actualisation est élevé. La période terminale T est appelée l'horizon temporel du calcul. Cet horizon peut être plus ou moins proche de la période présente ; tout dépend ici des hypothèses faites sur le caractère transitoire ou permanent des effets de la politique que l'on évalue, ou bien de la profondeur temporelle des bases de données statistiques permettant d'observer ces effets sur les revenus individuels (voir *infra*). Dans certains exercices de microsimulation, l'horizon temporel peut être l'âge de passage à la retraite, ou bien même l'âge potentiel de décès si l'on considère la variation de la pension de retraite induite par les effets de la politique considérée sur les revenus d'activité.

Si le montant moyen des coûts de mise en œuvre de cette politique pour un individu donné est égal à C , la valeur actualisée nette des revenus futurs de cet individu, lorsqu'il est bénéficiaire de cette politique, est égale à :

$$VAN_{[1,T]} = \left[\sum_{t=1}^T \frac{y_t}{(1+r)^t} - \frac{c_t}{(1+r)^t} \right]$$

Le coût c_t peut être éventuellement décomposé en un coût fixe c_{f0} et des coûts variables par période c_{vt} , de sorte que le coût par individu peut être écrit sous la forme :

$$C = C_0 + \sum_{t=1}^{T'} \frac{c_{vt}}{(1+r)^t}$$

Les coûts variables sont souvent encourus jusqu'à un horizon temporel T' plus rapproché que T , qui est celui qu'il faut prendre en compte pour le calcul des effets de la politique considérée sur les revenus individuels. L'exemple d'un effet de la réduction de la taille des classes développé plus bas en est une illustration.

Rappelons que le taux de rentabilité interne (TRI) est le taux d'actualisation qui annule la valeur actuelle nette ci-dessus définie, ou encore le taux qui égalise la somme des bénéfices futurs actualisés à celle des coûts futurs actualisés.

En France, le taux d'actualisation socioéconomique d'un investissement public préconisé par le rapport Quinet (2013)⁷⁰ est en moyenne de 4,5 %, notamment lorsque la sensibilité macroéconomique d'un projet est inconnue. Plus exactement, le rapport propose un taux d'actualisation sans risque de 2,5 % passant à 1,5 % au-delà de 2070. À ce taux, s'ajoute une prime de risque de 2 % pondérée par un coefficient spécifique à chaque projet en

⁶⁹ Dans les applications, la période de référence est souvent l'année.

⁷⁰ Rapport Quinet (2013), *L'évaluation socioéconomique des investissements publics*, tome 1.

fonction de la sensibilité de sa rentabilité à la croissance économique. Cette prime de risque dite systémique de 2 % est augmentée à 3 % pour les périodes d'évaluation au-delà de 2070⁷¹.

Nous recommandons dans ce guide d'utiliser ce taux dans l'évaluation socioéconomique de l'investissement social en France.

En matière d'investissement public, le taux d'actualisation choisi doit être le même pour toutes les alternatives envisagées et rester constant au cours du temps ; il doit permettre de valoriser suffisamment les périodes postérieures à la mise en œuvre de l'investissement réalisé, en raison notamment des intérêts des générations futures.

En ce sens, il s'oppose au taux d'actualisation privé, qui reflète la préférence pour le présent des personnes et qui peut dépendre de l'âge, du genre ou du niveau d'éducation des individus, mais aussi du contexte économique, comme l'ont montré plusieurs travaux de recherche exploitant les résultats d'expériences conduites auprès d'échantillons de taille et de composition diverses⁷². Le taux d'actualisation privé, tel qu'estimé par les chercheurs en économie expérimentale, est très supérieur à la valeur retenue dans les calibrations et simulations de modèles macro-économiques ou dans les analyses de type coût-bénéfices ; il est en outre très imprécis puisqu'il peut varier, selon les individus et les périodes considérées, entre 25 % et 35 %⁷³. Si l'on en croit ces travaux, les individus auraient donc une préférence pour le présent très élevée, et valoriseraient bien moins les périodes futures que la puissance publique, cela dénotant une certaine myopie des individus⁷⁴.

Mais l'investissement social, comme les autres formes d'investissements publics, concerne aussi des générations futures, et si l'État doit d'une certaine façon agréger les préférences, il doit aussi anticiper celles de ces générations à naître.

En conséquence, lorsqu'il s'agit d'évaluer les coûts et les bénéfices d'une politique d'investissement social supposée avoir des effets à court mais aussi à long terme, il est préférable de retenir le taux public de 4,5 % préconisé par le *Guide de l'évaluation socioéconomique des investissements publics* (2017). Les analyses de sensibilité menées devront tester d'autres valeurs de taux d'actualisation. Nous suggérons d'utiliser des taux variant de 3 % à 6 % pour ces tests.

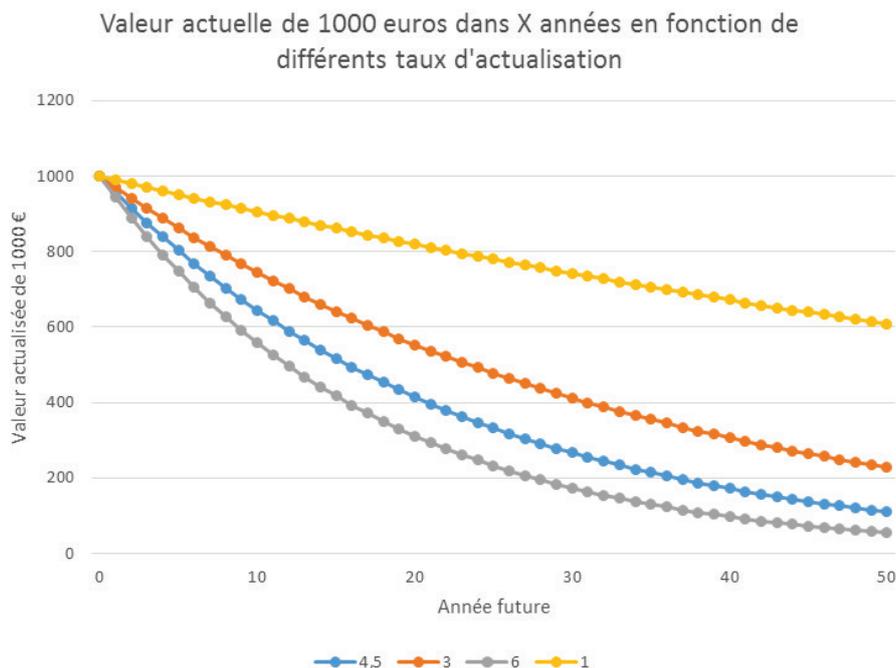
⁷¹ Voir les actes du colloque du 29 mars 2017 organisé par France Stratégie (2017) : *The discount rate in the evaluation of public investment project*, Proceedings of the conference, Commissariat général à l'investissement, Conseil général de l'environnement et du développement durable, France Stratégie, Paris.

⁷² Nous remercions vivement Marie-Claire Villeval qui nous a indiqué les principales références en ce domaine. Des synthèses récentes sont par exemple disponibles dans l'article de Cheung S. L. (2016), "Recent developments in the experimental elicitation of time preference", *Journal of Behavioral and Experimental Finance*, Vol. 11, p. 1-8 ; et dans Dhami S. (2016), "Behavioral time discounting", *The Foundations of Behavioral Economic Analysis*, chapitre III, Oxford University Press, novembre.

⁷³ Andreoni J. et Sprenger (C. 2012), "Estimating time preferences from convex budgets", *American Economic Review*, Vol. 102(7), p. 3333-3356.

⁷⁴ Le choix du taux d'actualisation n'est donc pas qu'un simple enjeu calculatoire : il revêt aussi un caractère éminemment politique.

Figure 6 – Illustration de la décroissance d'une valeur future en fonction du taux d'actualisation



2.2. Les revenus futurs

Les politiques d'investissement social, notamment celles qui ont pour but d'accroître le capital humain ou l'employabilité des personnes (qu'il s'agisse de politiques éducatives ou d'emploi), ont pour but d'améliorer l'insertion professionnelle et sociale des individus bénéficiaires de ces politiques. Il importe donc de mesurer cette variation induite par la politique publique, qu'il s'agisse de probabilités d'emploi (et donc de non-emploi) ou de qualité des emplois au cours des périodes futures. Une difficulté supplémentaire est de savoir si les effets de cette politique sont transitoires ou non, constants ou variables au cours du temps. Seules des études quantitatives menées à l'aide de données individuelles longitudinales permettent de répondre avec précision à cette question.

L'expression formelle de la VAN des effets de la politique dépend de ces paramètres. Notons respectivement $y_{(1)t}$ et $p_{(1)t}$ le revenu et la probabilité d'emploi d'un bénéficiaire de la politique considérée à la période future t . Si cet individu n'est pas employé à la période t , événement qui survient avec la probabilité $1 - p_{(1)t}$, son revenu de remplacement (allocation-chômage si la personne a cumulé des droits en ayant suffisamment travaillé auparavant, RSA sinon, etc.)⁷⁵ est égal à $b_{(1)t}$. Si le coût individuel de cette politique est égal à $c_{(1)}$, alors la VAN des flux de revenus futurs d'un individu bénéficiant de la politique est :

$$VAN_{(1),[1,T]} = \left[\sum_{t=1}^T \left(p_{(1)t} \frac{y_{(1),t}}{(1+r)^t} + (1 - p_{(1),t}) \frac{b_{(1),t}}{(1+r)^t} - \frac{c_{(1),t}}{(1+r)^t} \right) \right]$$

⁷⁵ Pour simplifier, on néglige ici les cas d'inactivité choisie.

Avec des notations similaires, la VAN des flux de revenus futurs de cet individu en l'absence de la politique a pour expression :

$$VAN_{(0),[1,T]} = \left[\sum_{t=1}^T \left(p_{(0),t} \frac{y_{(0),t}}{(1+r)^t} + (1 - p_{(0),t}) \frac{b_{(0),t}}{(1+r)^t} - \frac{c_{(0),t}}{(1+r)^t} \right) \right]$$

La politique d'emploi ou d'éducation considérée est supposée avoir des effets sur deux paramètres, le revenu $y_{(1),t}$ et la probabilité d'emploi $p_{(1),t}$ d'un individu à la période future t . Dans beaucoup d'analyses, le revenu est le salaire net perçu par l'individu, après déduction des cotisations sociales. Plus rarement, il s'agit d'un revenu net disponible après impôt. Le revenu ici pris en compte constitue donc un bénéfice « privé ». Il ne tient pas compte des transferts socio-fiscaux prélevés sur le salaire, qui constituent des bénéfices « publics ». En effet, si une politique publique d'emploi ou d'éducation a des effets positifs sur les revenus salariaux des individus qui en bénéficient, elle a également pour effet d'accroître les prélèvements socio-fiscaux (cotisations sociales, impôt sur le revenu, etc.) qui profitent à l'ensemble de la communauté⁷⁶. Pour intégrer les variations des contributions socio-fiscales, trois solutions peuvent être envisagées en fonction de la situation et des données disponibles.

Comment évaluer les variations de transferts socio-fiscaux (et gérer les données manquantes)

Tout d'abord, dans le cas d'une évaluation socioéconomique *ex post*, l'évolution des transferts socio-fiscaux peut être directement observée dans les données (administratives de préférence, ou par enquête auprès des ménages) et comparée entre les bénéficiaires et les non-bénéficiaires. C'est par exemple ce qui a été fait par Crépon *et al.* (2012)⁷⁷. Dans cette expérience aléatoire contrôlée menée en France au milieu des années 2000, des bénéficiaires du RMI ont été encouragés à participer à un programme d'accompagnement renforcé à la reprise d'activité et suivis pendant dix-huit mois (six trimestres). Dans leur analyse de l'impact de ce programme, les auteurs de cette étude identifient à la fois l'effet sur l'emploi et le montant de RMI versé aux personnes. Leur étude est assez singulière au regard des méthodes classiques d'analyse coûts-bénéfices puisqu'ils considèrent la valeur du programme au niveau individuel, comme la différence de coût du programme et la somme des écarts de RMI versés :

$$V_i = C_s^1 - \sum_{s \leq t < s+D} (T_t^0 - T_t^1)$$

où C_s^1 est le coût moyen du programme pour une personne entrée dans ce programme à la date s , T_t^0 le montant de RMI perçu dans la situation contrefactuelle (sans accompagnement) à la date t , et T_t^1 le montant de RMI perçu avec l'accompagnement à la même date. Cette valeur n'est pas actualisée et la valeur ne tient pas compte de l'effet sur l'emploi. La logique

⁷⁶ Les ensembles de barèmes collectés par l'Institut des politiques publiques (IPP, Paris), [disponibles à l'adresse suivante](#), permettent de calculer les montants de ces prélèvements pour un montant de salaire net donné.

⁷⁷ Crépon B., Gurgand M., Kamionka T. et Lequien L. (2012), "Is counseling welfare recipients cost effective? Lessons from a randomized experiment", Working Paper, J-Pal - Paris School of Economics. Cette étude a été présentée par Roland Rathelot devant le groupe de travail.

des auteurs est en accord avec celle du commanditaire de l'étude : à quelles conditions ce programme d'accompagnement est-il rentable, à quelle condition les économies de minima sociaux couvrent-elles le coût du programme ? La condition est la suivante :

$$E(C_s^1) \leq \sum_{s \leq t < s+D} (E(T_t^0) - E(T_t^1))$$

Dans cette expression, l'opérateur E représente l'espérance mathématique de la variable entre parenthèses. Les auteurs n'explicitent pas la relation entre le montant de RMI économisé et l'emploi, mais définissent la valeur implicite de l'emploi associé à ce programme. Celle-ci est définie par l'inégalité suivante :

$$W \left(\sum_{s \leq t < s+D} E(e_t^1) - E(e_t^0) \right) \geq E(C_s^1) - \sum_{s \leq t < s+D} (E(T_t^0) - E(T_t^1))$$

où $\sum_{s \leq t < s+D} E(e_t^1) - E(e_t^0)$ est l'effet moyen du programme sur l'emploi. À l'aide de cette expression, les auteurs peuvent soit identifier la valeur de W à partir de l'expérience, ou bien définir la valeur de W qui correspondrait à la disposition de la puissance publique à payer pour remettre ce groupe de personnes en activité (préférence révélée).

Si l'identification des effets est assurée par une méthodologie d'évaluation convaincante (comme une expérimentation aléatoire contrôlée par exemple), cette première méthode n'a besoin d'aucune autre hypothèse et permet d'identifier de façon très rigoureuse les variations de transferts induites par une politique. De façon tout à fait analogue, on pourrait observer l'évolution de l'emploi, des revenus, et ainsi estimer directement sans autre hypothèse les variations de revenus causées par l'investissement social étudié.

Cela suppose néanmoins de pouvoir appairer les données d'enquêtes aux données administratives socio-fiscales et de compléter les données disponibles. Dans le cas français par exemple, l'usage des Déclarations annuelles des données sociales (unifiées) (DADS et DADS-U) ou des Déclarations sociales nominatives (DSN) pour mesurer l'activité ou les revenus du travail ne tiennent pas (encore) compte de l'emploi public ou de l'emploi de particuliers.

La seconde méthode peut compléter ou se substituer à la première, mais elle suppose que l'analyste ait accès à des données individuelles. À ce titre, elle est davantage recommandée dans le cas d'évaluations *ex post*, mais elle peut être également adaptée à une évaluation *ex ante*. Il s'agit ici de simuler à l'aide d'un modèle de microsimulation les évolutions de transferts socio-fiscaux pour chaque individu, bénéficiaire ou non de la politique, et d'obtenir la somme des différences entre la situation après application de la politique et la situation contrefactuelle (voir section 3.IV pour une discussion sur les apports des modèles de microsimulation).

Cette méthode suppose que l'on dispose d'un modèle de microsimulation adapté, et elle impose de faire des hypothèses (par exemple, sur le taux de recours au RSA ou à la prime d'activité, la durée et les montants de droits au chômage, etc.) et de nombreuses

simplifications⁷⁸. La principale difficulté est la projection dans le temps car les modèles de microsimulation sont souvent statiques (*i. e.*, ils représentent le système socio-fiscal à l'instant présent), et parce qu'il faut alors faire « vieillir » l'échantillon, les montants de cotisation, etc.

Dès lors, on se rend rapidement compte qu'il devient difficile d'avoir une réponse précise et claire tant les hypothèses sont nombreuses. C'est pourquoi, le plus souvent, les variations de transferts socio-fiscaux sont estimées à partir de valeurs moyennes, globales mais cohérentes, et projetées à partir de données individuelles ou agrégées dans les analyses coûts bénéfiques. C'est ce qui est fait par Heckman et ses co-auteurs dans le cadre de l'analyse du programme Perry Preschool (cf. Annexe A)⁷⁹. De même, les variations de revenus annuels, de taux de salaire, de taux d'emplois, etc., ne sont souvent observées qu'en un nombre réduit d'occasions ; il est alors nécessaire de construire des trajectoires moyennes.

Enfin, dans bien des situations, l'analyste ne dispose d'aucune donnée sur les trajectoires de revenus car l'évaluation s'arrête trop tôt ou concerne des résultats indirects comme par exemple des résultats scolaires, la probabilité d'obtention d'un diplôme ou une mesure de l'état de santé d'une personne. Il est alors nécessaire de disposer d'une valeur monétaire pour ces résultats ou d'une matrice permettant de convertir les résultats.

Quelles valeurs tutélaires pour un accroissement du niveau éducatif ?

De nombreux travaux ont pu mesurer le lien entre augmentation du niveau d'éducation et revenus. Il s'agit d'un parti pris reflétant une littérature théorique et empirique abondante ayant démontré le lien causal entre investissement en éducation et revenu futur. En ce sens, une politique qui accroît le niveau d'éducation est censée augmenter à terme le revenu du bénéficiaire. Il s'agit donc d'un lien direct mais difficile à observer et à quantifier.

Dans la synthèse de référence qu'il a consacrée aux rendements salariaux de l'éducation, Card (2001)⁸⁰ trouve qu'en moyenne, et donc indépendamment du niveau d'éducation atteint, une année supplémentaire d'éducation est associée à un gain salarial compris entre 6 % et 10 % selon les époques et les pays ; si l'on suppose que les revenus de remplacement sont une proportion constante du dernier salaire et que les probabilités d'emploi ne sont pas significativement modifiées par une année d'éducation supplémentaire (hypothèses que l'on peut bien évidemment relâcher), l'estimation de Card consiste à borner le ratio des VAN :

⁷⁸ Par exemple, le RSA est doté d'un mécanisme dit de « cumul intégral » qui permet de cumuler RSA et salaire pendant un trimestre avant que celui-ci ne soit ajusté, et un mécanisme facultatif de « neutralisation des ressources » pour qu'en cas de perte d'emploi, le RSA puisse être rétabli avant le prochain trimestre d'actualisation des demandes. La simulation d'un tel mécanisme est très compliquée puisqu'elle dépend de la date de première inscription au RSA, de la date de reprise d'emploi par rapport au trimestre de référence, etc.

⁷⁹ Heckman J., Moon S. H., Pinto R., Savelyev P. et Yavitz A. (2010), "The rate of return to the High Scope Perry Preschool Program," *Journal of Public Economics*, Vol. 94(1-2), p. 114-128.

⁸⁰ Card D. (2001), "Estimating the return to schooling: Progress on some persistent econometric problems", *Econometrica*, Vol. 69(5), p. 1127-1160.

$$1,06 \leq \frac{VAN_{(1),[1,T]}}{VAN_{(0),[1,T]}} \leq 1,1$$

Ce qui correspond au ratio des VAN privées (*i.e.* ne tenant compte que de l'effet sur le revenu individuel et pas du différentiel de coût d'éducation).

Dans une étude fondée sur des comparaisons internationales, Maguain (2007)⁸¹ estime qu'en France le taux de rendement individuel d'une année supplémentaire d'éducation pour les hommes était égal à 7,5 % en 1995, valeur située dans l'intervalle identifié par Card (2001) ; selon Maguain (2007), ce taux est nettement plus élevé pour les années d'études supérieures, sa valeur étant d'environ 12 % ; mais cette dernière estimation ne tient pas compte des potentiels biais de sélection et d'auto-sélection qui déterminent la possibilité d'accéder à l'enseignement supérieur et d'y réussir. En clair, ce chiffrage ne permet pas de dire si les revenus plus élevés des plus diplômés sont dus à ce surcroît d'éducation ou au fait que les élèves les plus sûrs d'eux-mêmes prolongent leurs études quand ceux les plus en difficulté optent pour des études plus courtes. Caractériser l'ampleur de ce biais de sélection reste un défi pour les chercheurs comme le montre la récente étude de Kirkeboen *et al.* (2016)⁸².

À défaut d'estimation plus précise et causale, le rendement salarial d'une année d'étude supplémentaire peut être évalué entre 5 % et 10 %. Mais l'analyste qui souhaiterait utiliser ce chiffrage doit être conscient que cette estimation reflète en partie le phénomène d'auto-sélection ; l'existence d'un rendement privé moyen ne suffit pas à augmenter le taux d'inscription des étudiants dans l'enseignement supérieur.

Ce signal peut ne pas suffire à attirer les étudiants en cas de problèmes d'accessibilité financière aux études supérieures. Un exemple d'analyse coûts-bénéfices est contenu dans une étude française qui évalue les bourses sur critères sociaux. Celles-ci représentent le principal programme d'aides directes de l'État aux étudiants, tant du point de vue du nombre d'étudiants concernés que des montants distribués. Fack et Grenet (2015) ont évalué l'impact de ces bourses non seulement sur la décision d'entreprendre ou de poursuivre des études supérieures mais également sur la probabilité d'obtention d'un diplôme universitaire⁸³. Ces bourses, qui sont attribuées sous conditions de ressources, prennent en compte les revenus et les charges de l'étudiant et de sa famille (nombre de frères et sœurs, distance entre le domicile familial et le lieu d'études, handicap, etc.). Les résultats montrent que les bourses sur critères sociaux constituent un dispositif efficace pour améliorer les perspectives et les résultats universitaires des étudiants d'origine sociale modeste.

Fack et Grenet (2015) ont ensuite réalisé une analyse coûts-bénéfices de ce dispositif de bourses. Pour ce faire, ils ont comparé les coûts engendrés par l'offre d'une bourse annuelle de 1 500 euros à tous les bacheliers éligibles, et le bénéfice salarial qu'ils en tireraient s'ils

⁸¹ Maguain D. (2007), « Les rendements de l'éducation en comparaison internationale », *Économie et prévision*, n° 180-181, p. 87-106.

⁸² Kirkeboen L. J., Leuven E. et Mogstad M. (2016), "Field of Study, Earnings and Self-Selection", *Quarterly Journal of Economics*, 131(3), 1057-1111.

⁸³ Fack G. et Grenet J. (2015), "Improving College Access and Success for Low-Income Students: Evidence from a Large Need-Based Grant Program", *American Economic Journal: Applied Economics*, Vol.7(2), p. 1-34. Gabrielle Fack a présenté cette étude devant le groupe de travail.

obtenaient une licence ou un master. Les coûts de cette mesure incluent 1) le montant total des bourses offertes à tous les étudiants éligibles ; 2) le coût de formation de ces étudiants lorsqu'ils s'inscrivent au moins pendant deux ans à l'université à la suite de l'attribution de la bourse⁸⁴ ; et 3) le coût d'opportunité associé aux deux années pendant lesquelles ces étudiants boursiers renoncent à travailler (à temps plein) comme ils devraient le faire en arrêtant leurs études après le baccalauréat. Le coût total associé à l'offre d'une bourse annuelle de 1 500 euros à tous les étudiants éligibles serait alors de 100 000 euros pour l'obtention d'une licence, et de 125 000 euros pour celle d'un master. Le bénéfice salarial d'une formation universitaire est calculé à partir des salaires perçus par les individus âgés de 18 à 60 ans dans l'enquête sur l'emploi de 2011. Les salaires sont supposés croître de 1 % par an. Fack et Grenet ont alors estimé qu'en valeur actualisée, un étudiant titulaire d'une licence percevrait en moyenne 145 000 euros de plus qu'un bachelier durant toute sa vie active, et qu'un titulaire de master percevrait 425 000 euros de plus. Ces flux de salaires actualisés sont supérieurs aux coûts de la mesure envisagée. Dans tous leurs calculs, Fack et Grenet (2015) ont fait l'hypothèse que le taux d'actualisation est égal à 3 %⁸⁵.

France Stratégie⁸⁶ a produit en 2019 un rapport proposant un cadre pour l'évaluation socioéconomique de projets immobiliers dans l'enseignement supérieur. Ce type d'investissements relève davantage du modèle classique des grands investissements publics. À leur différence, de tels investissements ont néanmoins pour objectif d'accroître le capital humain des étudiants qui en bénéficient, et en ce sens, ils relèvent du champ de l'investissement social. Les auteurs du rapport s'appuient sur les analyses de plusieurs groupes de travail qui ont défini des valeurs tutélaires pour un certain nombre de variables de résultat. En particulier, les gains privés liés à l'obtention d'un diplôme sont obtenus à partir de modèles économétriques à la « Mincer » semblables à ceux présentés plus tôt. Ces modèles font l'hypothèse qu'il n'y a pas d'effet de sélection à l'entrée dans l'enseignement supérieur et que les différences de trajectoires de revenus correspondant aux niveaux et filières sont les bénéfices directs de l'investissement en éducation. Ces hypothèses sont bien entendu très fortes. Cependant, l'intérêt du guide produit par France Stratégie est de recommander l'utilisation de valeurs tutélaires de façon à ce qu'en comparant des projets, « l'erreur » soit constante dans les différentes évaluations et que, par différence de valeurs actualisées, les erreurs systématiques s'annulent.

Ainsi, à défaut d'observation directe, l'effet sur les revenus de l'allongement d'une année de la durée de scolarité peut être approximé par une augmentation moyenne du revenu de 5 % à 10 %. La valeur de l'obtention d'un diplôme est fournie par le rapport de France Stratégie.

⁸⁴ Selon le ministère de l'Éducation, le coût moyen d'un étudiant inscrit à l'Université était de 10 770 euros en 2011. Avec un taux d'actualisation de 3 %, le coût de ses deux premières années universitaires est donc estimé à 21 200 euros.

⁸⁵ Ce coût d'opportunité est estimé grâce aux données de l'enquête Emploi de l'Insee qui permet d'observer les salaires des diplômés de l'enseignement supérieur en 2011. Selon ces données, le coût d'opportunité de deux années d'études supérieures serait de 40 000 euros pour un étudiant de licence et de 50 000 euros pour un étudiant de master.

⁸⁶ France Stratégie (2019), *L'évaluation socioéconomique des projets immobiliers dans l'enseignement supérieur et la recherche*, sous la direction d'E. Quinet, rapporteur, J. Ni., Paris.

Politiques actives et revenus : quels effets pour quelles valeurs ?

La plupart des politiques actives d'emploi (ALMP ou *active labour market policies*), qu'il s'agisse de subventions à l'embauche (telles que les réductions de cotisations patronales ou les crédits d'impôt) ou de programmes de formation continue destinés aux salariés ou aux chômeurs, ont principalement pour objectif d'améliorer les perspectives d'emploi des travailleurs non qualifiés. La valeur sociale d'un tel programme est difficile à appréhender car elle dépend des hypothèses sur le mécanisme en jeu : amélioration de la qualité de l'appariement, augmentation du taux de mise en contact avec les emplois vacants, déplacement dans la file d'attente, acquisition de compétences nouvelles et de capital humain, etc. Suivant les cas, l'effet sur les revenus peut être transitoire, durable, ou totalement absent dans le cas d'un déplacement dans la file d'attente.

De ce fait, peu d'études ont examiné les effets des politiques actives d'emploi sur les salaires. La plupart ont estimé l'effet sur la probabilité d'emploi après participation au programme, probabilité qui intervient dans la formule précédente de la VAN.

Dans une méta-analyse récente, Card, Kluve et Weber (2018)⁸⁷ ont passé en revue 207 études consacrées à cinq types de politiques actives d'emploi : les programmes de formation *stricto sensu*, les aides à la recherche d'emploi, les sanctions appliquées aux salariés ou aux chômeurs refusant de participer aux programmes de formation ou d'accepter des offres d'emploi conformes à leur qualification, les subventions à l'emploi relatives aux entreprises du secteur privé, et les emplois aidés du secteur public⁸⁸. Ces 207 études fournissent 857 estimations d'effets pour un total de 528 groupes de salariés ayant participé aux programmes recensés. Cette méta-analyse est donc l'une des plus complètes dont nous disposons à ce jour. Les pays les plus représentés dans cette analyse sont l'Allemagne (53 estimations), le Danemark (115 estimations), la Suède (66 estimations), les U.S.A. (57 estimations) et la France (42 estimations).

Card, Kluve et Weber (2018) ont classifié les résultats en fonction des types de programmes et des groupes « cible » de personnes éligibles. Ils ont distingué par la suite trois horizons temporels : le court terme (un an après la participation au programme), le moyen terme (un à deux ans après) et le long terme (plus de deux ans après). Ils retiennent pour chaque intervention ou programme considéré la taille de son effet (*effect size*) et sa significativité statistique⁸⁹. En raison des groupes ciblés, composés de personnes peu qualifiées ou de chômeurs de longue durée, ils privilégient comme c'est classiquement le cas les effets sur la probabilité d'être en emploi un an, deux ans ou plus

⁸⁷ Card D., Kluve J. et Weber A. (2018), "What Works? A Meta-Analysis of Recent Active Labor Market Program Evaluations", *Journal of the European Economic Association*, Vol 16(3), p. 894-931.

⁸⁸ Les mesures relatives à la baisse du montant et de la durée des allocations-chômage ne sont pas des politiques actives d'emploi. En d'autres termes ce ne sont pas des politiques d'investissement social. C'est pour cette raison qu'elles ne sont pas citées ici.

⁸⁹ La taille de l'effet du programme est obtenue comme son effet en valeur absolue divisé par l'écart-type de la variable de résultat (par exemple, le fait d'être en emploi) dans le groupe de contrôle. Cette mesure est préférée à la significativité statistique de l'effet estimé car elle ne dépend pas de la taille de l'échantillon. Les deux mesures sont toutefois complémentaires et ont le même signe, positif ou négatif. Card, Kluve et Weber (2018) retiennent trois catégories pour caractériser la significativité statistique de l'effet estimé : significativement positif, non significativement différent de zéro, et significativement négatif.

de deux ans après participation au programme. Les conclusions principales de leur étude sont les suivantes :

- **les effets moyens des programmes d'activation sont en général très faibles dans le court terme, mais plus élevés à moyen et long termes** ; la taille moyenne de l'effet sur l'emploi un an après participation à un programme de ce type est statistiquement nulle, mais elle augmente pour devenir égale à 0,12 écart-type l'année suivante, puis à 0,19 écart-type la troisième année, ces deux dernières estimations étant statistiquement significatives ;
- **cette progression temporelle varie en fonction du type de programme** ;
- **les effets des programmes d'activation varient d'un groupe cible à l'autre**. Certains types de programmes sont plus efficaces lorsqu'ils sont offerts à certains groupes de personnes ;
- les programmes d'activation sont **plus efficaces dans les périodes de récession que dans les phases hautes du cycle** où le nombre de créations d'emploi est plus élevé.

La mise en œuvre de programmes de placement ou de formation des chômeurs peut toutefois se heurter à des effets de déplacement, tels que Crépon *et al.* (2013)⁹⁰ les ont mis en évidence. L'effet ici étudié est celui d'un programme d'aide au placement de jeunes chômeurs diplômés, qui est évalué à l'aide d'une expérimentation randomisée conduite à grande échelle sur l'ensemble du territoire français. Ce programme se révèle efficace à court terme puisque, huit mois après leur participation au programme, les jeunes chômeurs sont plus souvent en emploi stable que ceux, membres du groupe de contrôle, n'en ayant pas bénéficié. Mais ce bénéfice n'est pas persistant. Il se fait par ailleurs au détriment des chômeurs non bénéficiaires. Pour le dire autrement, plus la proportion de jeunes chômeurs bénéficiaires augmente, plus ceux qui sont aléatoirement écartés du programme trouvent difficilement un emploi. Ce phénomène s'observe évidemment plus souvent dans les bassins d'emploi où le nombre de créations d'emploi est faible.

Au total donc, le programme de placement évalué a peu d'effets positifs. Ce résultat est conforme à l'une des conclusions de la méta-analyse de Card, Kluve et Weber (2018), à savoir que les programmes d'aide à la recherche d'emploi n'ont que peu ou pas d'effet sur le devenir des chômeurs à court et plus long termes.

La littérature académique ne fournit quasiment pas d'estimations quantifiées des effets des politiques d'activation sur les revenus. En l'état actuel des connaissances, des études complémentaires doivent être réalisées en vue d'estimer de façon robuste les effets des politiques d'activation sur le revenu.

⁹⁰ Crépon B., Duflo E., Gurgand M., Rathelot R. et Zamora P. (2013), "Do Labor Market Policies have Displacement Effects? Evidence from a Clustered Randomized Experiment", *The Quarterly Journal of Economics*, Vol. 128(2), p. 531-580. Roland Rathelot a présenté cette étude devant le groupe de travail.

Que retenir ?

Le taux d'actualisation permet de ramener à une même date des flux monétaires intervenant à des années différentes, et donc de faire l'arbitrage entre le présent et le futur.

En France, le taux d'actualisation socioéconomique d'un investissement public préconisé par le rapport Quinet (2013) est de 4,5 %, valeur qu'il convient de reprendre dans les analyses socioéconomiques.

Les politiques d'investissement social ont pour but d'accroître le capital humain et, par conséquent, d'améliorer l'insertion professionnelle et sociale des individus.

Dans le calcul de leur VAN, au-delà des conséquences sur les revenus futurs, il faut aussi tenir compte des effets sur les probabilités d'emploi qui peuvent varier dans le temps.

Lorsqu'il améliore l'employabilité des individus, l'investissement social a également pour effet de modifier les prélèvements socio-fiscaux (cotisations sociales, impôt sur le revenu, etc.) et les revenus de substitution (allocations chômage, RSA, etc.). Il faut de ce fait les intégrer dans la VAN en tant que bénéfiques pour la société, et les distinguer des bénéfiques individuels.

La valeur d'un allongement de la durée des études sur les revenus futurs peut être estimée entre 5 % et 15 % par an, la valeur de l'obtention d'un diplôme est fournie par le rapport de France Stratégie publié en 2019.

La littérature sur les politiques actives d'emploi concerne principalement les effets sur la probabilité d'emploi, mais l'effet sur les revenus futurs est très peu étudié.

3. Comment convertir des effets intermédiaires et prendre en compte les effets inobservés (connus)?

Les valeurs tutélaires proposées avec précaution dans la section précédente concernent les effets d'un allongement de la durée de la scolarité ou de l'obtention d'un diplôme sur les revenus futurs. Cependant, ces valeurs ne représentent pas tous les effets d'un investissement dans l'éducation. En effet, les politiques éducatives ont souvent pour objectif d'intervenir sur d'autres dimensions du développement personnel ou sur des objectifs spécifiques.

De fait, les interventions existantes ont souvent des résultats qui ne sont pas la durée de la scolarité ou l'obtention d'un diplôme. Par ailleurs, beaucoup de ces résultats sont liés entre eux et ne sont pas nécessairement estimés dans le cadre des évaluations conduites. La suite de cette section illustre le premier enjeu en décrivant les méthodes utilisées pour « convertir » des progrès scolaires en revenus futurs. Elle présente ensuite la façon de tenir compte de résultats intermédiaires.

3.1. Le problème des résultats scolaires et du taux de salaire

Lorsque l'on examine l'efficacité d'une politique éducative (par exemple, une réduction de la taille des classes) dont les effets à court et moyen termes sont généralement mesurés en unités d'écart-type de la distribution aux résultats des examens ou tests scolaires, une question centrale est de savoir comment passer de cette mesure à celle des effets de la politique sur l'insertion professionnelle et les revenus perçus dans le futur (au moins dans les premières années de vie active). En d'autres termes, de combien croissent les revenus futurs lorsque les résultats scolaires en fin de cycle secondaire (mesurés par des tests standardisés ou des examens nationaux) augmentent d'une certaine quantité, celle-ci étant généralement mesurée en unités d'écart-type de la distribution des notes à ces tests ? Cette échelle de mesure est clairement expliquée dans une note de l'Institut des politiques publiques : « L'effet d'une politique éducative sur la performance des élèves est souvent mesuré à partir de tests soumis aux élèves. Ces tests n'ont pas d'échelle naturelle, ils peuvent être plus ou moins difficiles, compter plus ou moins de questions, etc., et leur valeur absolue n'a pas une signification précise. Il est d'usage dans la littérature de recherche de les rendre comparables en les normalisant par l'écart-type du test dans la population. Techniquement, il s'agit de diviser la note obtenue par chaque élève par l'écart-type de l'échantillon [...]. Lorsque les résultats sont normalisés par leur écart-type, ils expriment davantage un niveau relatif qu'une performance absolue. Une fois normalisés, tous les tests ont un écart-type ramené à 1. Alors, si l'effet d'une intervention est d'augmenter le test d'une unité (ce qui signifie un écart-type, ou 100 % d'un écart-type), cela revient à amener un élève médian (le 12^e sur 24 élèves, par exemple) au niveau de l'élève qui est au 4^e rang sur 24, soit une progression de 8 rangs dans la classe. Un effet de 50 % d'un écart-type correspond, pour l'élève médian, à une progression de 4 rangs dans une classe de 24⁹¹ . »

Certaines études proposent des valeurs pour ce rapport entre résultats scolaires et revenus futurs. Les valeurs obtenues divergent sensiblement d'un pays ou d'une période à l'autre, de sorte qu'il est impératif d'obtenir à l'avenir des mesures plus précises.

Par exemple, Murnane, Willet et Levy (1995)⁹² et Neal et Johnson (1996)⁹³ ont trouvé **qu'une variation d'un écart-type des résultats en fin de lycée est associée à une augmentation des revenus professionnels d'environ 8 % onze ans plus tard**. Cette valeur est utilisée par Krueger et Whitmore (2001)⁹⁴ dans leur analyse coûts-bénéfices de la réduction de la taille des classes prévue par le programme STAR aux États-Unis (cf. sous-section 4.1). Filatriau, Fougère et Tô (2013)⁹⁵ estiment pour leur part, à partir de l'enquête « *Entrée dans la vie active* » conduite par le ministère de l'Éducation et l'Insee, **qu'une augmentation d'un écart-type des notes obtenues aux tests de mathématiques à l'entrée en sixième est associée à un**

⁹¹ Bouguen A., Grenet J. et Gurgand M. (2017), « La taille des classes influence-t-elle la réussite scolaire ? », Note n° 28 de l'I.P.P., Paris.

⁹² Murnane R., Willett J. et Levy F. (1995), "The Growing Importance of Cognitive Skills in Wage Determination", *The Review of Economics and Statistics*, Vol. 77(2), p. 251-66.

⁹³ Neal D. et Johnson W. (1996), "The Role of Premarket Factors in Black-White Wage Differences", *Journal of Political Economy*, Vol 104(5), p. 869-895.

⁹⁴ Krueger A. et Whitmore D. (2001), "The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR", *The Economic Journal*, Vol. 111(468), p. 1-28.

⁹⁵ Filatriau O., Fougère D. et Tô M. (2013), "Will Sooner Be Better? The Impact of Early Preschool Enrollment on Cognitive and Noncognitive Achievement of Children", Discussion Paper, No 9480, C.E.P.R., Londres.

accroissement du revenu professionnel de 12,5 % trois ans après l'entrée dans la vie active, alors qu'une augmentation équivalente des résultats aux tests de français ne produit qu'un accroissement de 2 % de ces revenus.

Machin et McNally (2008) ont évalué un programme appelé "literacy hour" (littéralement, l'heure de littératie) qui a été introduit durant l'année scolaire 1998-1999 dans les écoles élémentaires anglaises⁹⁶. Ce programme avait principalement pour but de modifier, en les unifiant, les pratiques pédagogiques des enseignants des écoles élémentaires en matière d'apprentissage de la lecture, de l'orthographe et de la grammaire (de manière plus générale, en matière de littératie définie par l'OCDE comme « l'aptitude à comprendre et à utiliser l'information écrite dans la vie courante, à la maison, au travail et dans la collectivité en vue d'atteindre des buts personnels et d'étendre ses connaissances et ses capacités »). Le programme s'adressait aux élèves de 5 à 11 ans. Il préconisait de modifier l'emploi du temps de chaque classe de sorte à garantir aux élèves une heure consacrée à la littératie chaque jour. Cette heure devait être divisée en trois temps : 10 à 15 minutes consacrées à la lecture et à l'écriture avec toute la classe ; 10 à 15 minutes consacrées au travail sur les mots (phonétique, orthographe et vocabulaire) et sur les phrases (grammaire et ponctuation), toujours avec la classe entière ; 25 à 30 minutes de travaux de groupes sur l'écriture et la lecture ; et enfin 5 à 10 minutes permettant de résumer les objectifs de la leçon, son apport, et de prévoir le contenu des leçons suivantes. Machin et McNally (2008) estiment que ce programme a permis d'améliorer les compétences des élèves en lecture et en compétences linguistiques de 2 % à 3 %. Conformément à cet effet relativement faible, **ils en ont déduit qu'une amélioration du score en lecture de 0,09 unités d'écart-type à l'âge de 10 ans a pour effet de faire croître les salaires à l'âge de 30 ans d'un pourcentage compris entre 0,5 et 1,4 %**. À supposer que l'effet soit linéaire, ceci impliquerait qu'un accroissement d'un écart-type correspondrait à une augmentation de salaire comprise entre 5,5 % et 15,5 %. Pour réaliser cette estimation, les auteurs utilisent d'autres sources de données relatives aux effets des évaluations scolaires en primaire sur les résultats en fin de lycée et sur les salaires à 30 ans.

Lorsque l'on mène l'analyse coûts-bénéfices de l'effet d'une politique éducative jugée efficace, les résultats des études précitées nous conduisent donc à proposer d'associer une croissance de 5 % à 15 % des revenus professionnels futurs à une amélioration des résultats scolaires d'un écart-type. Pour une amélioration inférieure, nous proposons d'appliquer une interpolation linéaire⁹⁷, ce qui revient à postuler des rendements constants du surcroît de réussite : par exemple, 2,5 % à 7,5 % pour un accroissement d'un demi-écart-type.

3.2. Prise en compte d'effets liés dans les analyses coûts bénéfiques

Dans de nombreux cas, les effets des politiques d'investissement social portent sur des résultats eux-mêmes liés à d'autres effets. Par exemple, l'évaluation d'un programme éducatif destiné à de jeunes enfants peut permettre d'observer directement leur propension à avoir un comportement perturbateur en classe, mais pas leur probabilité d'obtenir un diplôme de fin de second

⁹⁶ Machin S. et McNally S. (2008), "The literacy hour", *Journal of Public Economics*, vol. 92(5-6), p. 1441-1462.

⁹⁷ L'interpolation linéaire est la méthode la plus simple pour estimer la valeur prise par une fonction continue entre deux points déterminés. Elle consiste à utiliser pour cela la droite passant par les deux points déterminés.

cycle, d'être scolarisé dans une filière éducative adaptée, ou de commettre des infractions. Cependant, il existe d'autres sources liant les comportements perturbateurs à ces résultats.

Ne pas prendre en compte ces effets qui ne sont pas directement observés conduirait à sous-estimer les véritables effets de la politique d'investissement social. Leur prise en compte doit se fonder sur une connaissance théorique et empirique complète pour ne pas omettre ou ajouter de relations fallacieuses. La définition des résultats à prendre en compte et des conséquences pour l'évaluation est détaillée dans la section 3.VII. Une façon de procéder consiste à avoir recours à une revue systématique de la littérature et à synthétiser les résultats des estimations antérieures dans une méta-analyse selon les mêmes techniques et critères que ceux présentés dans le chapitre 2.IV. À partir de ces résultats, il est alors possible d'extrapoler les estimations sur une plus longue période. Il faut toutefois faire attention à ne pas additionner les effets croisés sans recourir à une structure de causalité. Les effets croisés doivent être recensés de façon systématique, en suivant un plan de préanalyse de façon à ne pas sélectionner uniquement les effets favorables. La crédibilité de cette démarche réside dans sa dimension systématique. Le Washington State Institute for Public Policy (WSIPP ci-après) utilise cette méthode pour lier les résultats comme le montre le Tableau 2.

Le WSIPP croise ainsi l'effet de la consommation de cannabis avant 18 ans à la probabilité de commettre des délits, d'obtenir un diplôme du secondaire. Il renseigne sur les effets des négligences et des abus dont les enfants sont victimes sur la consommation d'alcool, les troubles de l'anxiété, la probabilité de redoubler, d'être déprimé, de commettre un délit, etc. Par exemple, le fait pour un enfant d'avoir été abusé ou négligé augmente de 54 % d'un écart-type la probabilité de commettre un délit à 20 ans. Ce résultat est estimé à partir d'un ensemble de onze études.

Tableau 2 – Méta-analyse des liens entre diverses interventions et différentes mesures de résultats

Estimated causal links between outcomes	No. of effect sizes	Meta-analytic results before adjusting effect sizes									Adjusted effect size and standard error used in the benefit-cost analysis		Age of link measurement	Age at which relationship begins
		Fixed effects model			Random effects									
		Weighted mean effect size & p-value			Homogeneity test (p-value to reject homogeneity)			Weighted mean effect size & p-value			ES	SE		
ES	SE	p-value	Q-stat	p-value	ES	SE	p-value	ES	SE					
High school graduation	13	-0.180	0.016	0.000	109.830	0.000	-0.235	0.064	0.000	-0.235	0.064	18	18	
Cannabis use < 18 years of age, leading to...														
Crime	See cannabis use < 14 years of age (note: cannabis use < 18 has unique age references)											20	15	
High school graduation	See cannabis use < 14 years of age													
Cesarean section, leading to...														
Hospital readmissions	1	0.379	0.010	0.000	0.000	0.000	0.379	0.010	0.000	0.379	0.010	25	1	
Child abuse & neglect, leading to...														
Alcohol (disordered use)	6	0.171	0.028	0.000	7.590	0.180	0.172	0.046	0.000	0.172	0.046	25	18	
Anxiety (incl. OCD)	3	0.298	0.052	0.000	17.366	0.000	0.325	0.166	0.051	0.325	0.166	20	18	
Crime	11	0.532	0.034	0.000	35.330	0.000	0.542	0.071	0.000	0.542	0.071	20	18	
Depression	8	0.305	0.028	0.000	22.675	0.002	0.293	0.058	0.000	0.293	0.058	20	18	
Disruptive behavior	1	0.460	0.391	0.239	0.000	0.000	0.460	0.391	0.239	0.460	0.391	12	12	
Grade retention	1	0.446	0.102	0.000	0.000	0.000	0.446	0.102	0.000	0.446	0.102	12	17	
High school graduation	5	-0.412	0.048	0.000	14.308	0.006	-0.405	0.098	0.000	-0.405	0.098	18	18	
Illicit drugs (disordered use)	6	0.241	0.042	0.000	11.772	0.038	0.268	0.069	0.000	0.268	0.069	21	18	
Obesity	5	0.022	0.018	0.242	9.052	0.060	0.042	0.039	0.283	0.042	0.039	35	18	
PTSD	1	0.836	0.199	0.000	0.000	0.000	0.836	0.199	0.000	0.836	0.199	18	18	
Special education	1	0.389	0.036	0.000	0.000	0.000	0.389	0.036	0.000	0.389	0.036	8	5	
Test scores-academic	2	-0.270	0.062	0.000	2.278	0.320	-0.268	0.067	0.000	-0.268	0.067	17	17	
Tobacco (regular use)	1	0.387	0.123	0.002	0.000	0.000	0.387	0.123	0.002	0.387	0.123	20	18	
Crime (non-offender pop), leading to...														
High school graduation	6	-0.421	0.029	0.000	23.957	0.000	-0.505	0.079	0.000	-0.505	0.079	18	18	
Crime (offender pop), leading to...														
High school graduation	4	-0.174	0.043	0.000	6.516	0.089	-0.191	0.066	0.004	-0.191	0.066	18	18	

Source : WSIPP (2018)

Que retenir ?

Lorsque l'on examine l'efficacité d'une politique éducative dont les effets sont généralement mesurés en termes d'unités d'écart-type aux résultats des tests scolaires, une question centrale est de savoir comment passer de cette mesure à celle des effets sur les revenus perçus.

Les résultats des études conduisent à proposer d'associer une croissance de 5 % à 15 % des revenus professionnels futurs à une amélioration des résultats scolaires d'un écart-type. Pour une amélioration inférieure, nous proposons d'appliquer une interpolation linéaire.

Beaucoup de politiques d'investissement social affectent différentes dimensions du développement personnel, et ces résultats intermédiaires peuvent n'être que partiellement observés. Il convient donc de procéder à une revue systématique de littérature pour mesurer les effets liés et pouvoir projeter les effets au-delà de ce que l'évaluation permet.

4. Quel horizon temporel et quelles données privilégier ?

L'horizon temporel qui est pris en compte dans le calcul de la VAN doit être défini en fonction de la politique publique considérée. Il est différent selon que l'on s'intéresse à des politiques éducatives agissant sur des élèves du cycle primaire ou à des politiques d'emploi proposant des formations à des chômeurs adultes de longue durée. Il peut également dépendre du caractère permanent des effets anticipés de la politique considérée.

Dans l'idéal, on voudrait pouvoir disposer d'un suivi des bénéficiaires et de groupe(s) de comparaisons valides sur une longue période, mais bien souvent on doit se contenter d'utiliser d'autres données que l'observation des individus directement affectés.

Pour des raisons assez évidentes, il est difficile de suivre les bénéficiaires d'une politique sociale ou éducative pendant plusieurs années, à moins de pouvoir les repérer dans les fichiers administratifs ou dans ceux de la statistique publique à l'aide d'un identifiant et des autorisations nécessaires. De ce fait, pour réaliser l'analyse coûts-bénéfices d'une politique sociale, il faut souvent disposer de deux sources d'information : la première provient de l'observation directe des effets de l'intervention pour un échantillon de bénéficiaires et de personnes du groupe témoin, la seconde consiste en des fichiers longitudinaux qui permettent d'identifier des personnes comparables aux bénéficiaires et pour lesquelles on peut disposer d'informations pertinentes pour la réalisation du calcul coûts-bénéfices.

En France, plusieurs de ces sources existent déjà et pourraient être utilisées. On peut citer par exemple les fichiers d'élèves que produisent les services statistiques du ministère de l'Éducation nationale ou du ministère de l'Enseignement supérieur et de la Recherche, des enquêtes sur l'emploi réalisées par l'Insee, le fichier historique de Pôle Emploi, l'échantillon démographique permanent (EDP), les fichiers de déclarations annuelles de données sociales (DADS), les cohortes Elfe ou Constances gérées par l'Ined et/ou l'Inserm, etc. De tels fichiers fournissent des informations sur les parcours des individus en fonction de leur âge, de leur genre, de leur niveau d'éducation, de leur lieu de naissance, etc. Ces données

pourraient être complétées. Ces différentes sources de données peuvent être associées de façon à simuler les trajectoires futures des bénéficiaires et des non-bénéficiaires en matière de revenus, d'emploi, de santé, de conditions de vie, etc., comme le font Garcia, Heckman et al. (2017) dans une analyse exemplaire des bénéfices et des coûts de long terme de deux programmes éducatifs destinés aux très jeunes enfants, le Carolina Abecedarian Project et le Carolina Approach to Responsive Education (CARE).

Idéalement, ces appariements de fichiers devraient permettre de simuler les trajectoires individuelles pendant plusieurs dizaines d'années. C'est ce que font habituellement les outils de microsimulation qui, en ce domaine, constituent des compléments méthodologiques importants⁹⁸. Selon Didier Blanchet, la microsimulation « partage avec les approches expérimentales et pseudo-expérimentales le fait d'accorder une attention particulière aux questions d'hétérogénéité individuelle ». Mais, à la différence de ces approches, **la microsimulation permet d'apprécier quantitativement les effets redistributifs et pourrait permettre la prise en compte de certains effets de bouclage, autrement dit les effets qu'une politique publique peut avoir sur d'autres groupes que les groupes de traitement et de contrôle utilisés dans le cadre de l'analyse expérimentale ou quasi expérimentale**⁹⁹. Ces effets de bouclage incluent les contraintes et conséquences budgétaires d'une politique publique, ainsi que les transferts sociaux qu'elle peut engendrer. Une fois les effets de la politique microsimulés, on peut ainsi les analyser sous des angles très variés : agrégats ou moyennes sur des sous-populations, indicateurs de dispersion, indicateurs de bilans intertemporels dans le cas de modèles dynamiques, etc.¹⁰⁰

S'agissant des politiques d'emploi ou de santé, que celles-ci soient évaluées au moyen d'expérimentations randomisées ou à l'aide d'observations quasi expérimentales, les durées d'observation des groupes cible et témoin devraient être aussi longues que possible. L'objectif est ici de pouvoir mesurer les effets de moyen/long terme de politiques dont on sait que certaines n'ont pas d'effet à court terme, alors que d'autres voient leurs effets significatifs à court terme s'effacer très rapidement dans les deux ou trois années qui suivent. L'observation directe des cohortes est ici essentielle.

Évidemment, la situation idéale serait, plutôt que de procéder à des appariements de fichiers, de suivre directement et sur le long terme les individus pour lesquels les évaluations sont menées. Dans le cas des politiques éducatives (réduction de la taille des classes, utilisation d'outils numériques pour l'enseignement, etc.), le suivi des bénéficiaires devrait ainsi être effectué a minima jusqu'à l'étape d'entrée sur le marché du travail, voire cinq à dix

⁹⁸ Sur la complémentarité entre l'approche expérimentale ou quasi expérimentale et les méthodes de microsimulation, on se référera à l'article de Blanchet D., Hagneré C., Legendre F. et Thibault F. (2016), « Évaluation des politiques publiques, *ex post* et *ex ante* : l'apport de la microsimulation », *Revue économique*, Vol. 67(4), p. 685-696.

⁹⁹ Il faut toutefois émettre ici une mise en garde : prétendre tout simuler exposerait au « risque modélisateur » préalablement évoqué.

¹⁰⁰ Didier Blanchet a exposé ces arguments devant le groupe de travail dans le cadre d'une communication intitulée « Modèles de microsimulation : quelle complémentarité avec les autres outils d'évaluation des politiques sociales ? ».

ans après l'entrée dans la vie active, à l'instar de l'étude de Chetty et al. (2011)¹⁰¹ ou de celle de Fredriksson et al. (2013)¹⁰² sur les effets à long terme de la réduction de la taille des classes. L'observation devrait également permettre de mesurer la fluidité du parcours éducatif préalable (probabilités de redoubler ou d'accéder à l'enseignement supérieur).

Que retenir ?

L'horizon temporel qui est pris en compte dans le calcul de la VAN doit être défini en fonction de la politique publique considérée.

Pour y parvenir, il est nécessaire de suivre les populations concernées à partir d'enquêtes longitudinales et/ou de procéder à des appariements de fichiers statistiques.

Ces appariements de fichiers devraient permettre de simuler les trajectoires individuelles pendant plusieurs dizaines d'années. C'est ce que font habituellement les outils de micro-simulation qui permettent en particulier d'estimer les effets redistributifs et de prendre en compte certains effets de bouclage.

5. Le ratio coûts-efficacité

Parfois, notamment lorsqu'il s'agit d'évaluer des politiques d'emploi ou de santé qui ont un objectif principal à court terme et dont les effets attendus à long terme sont faibles, les analystes préfèrent utiliser le ratio coût-efficacité, plutôt que la valeur actualisée nette dont la formule est plus complexe et repose sur des composantes souvent plus aléatoires. Le ratio coûts-efficacité est défini comme le coût par unité d'effet obtenu. Des exemples simples sont le coût par emploi créé dans le cas d'une politique de réduction des charges salariales, ou le coût par vie sauvée dans celui d'une nouvelle thérapie. Généralement, ce type de calcul se fait sur un horizon de temps plus court, c'est-à-dire en ne tenant compte que des coûts et de l'efficacité du programme évalué tels qu'ils peuvent être mesurés durant les périodes qui suivent son introduction. Une fois l'objectif du programme clairement défini et identifié (par exemple, le nombre d'emplois créés durant les trois années de mise en œuvre du dispositif) et le coût total du programme évalué (obtenu en sommant toutes les dépenses publiques qui ont été nécessaires pour obtenir le résultat recherché), il suffit de calculer le ratio :

$$\text{Coût total} / \text{Mesure de l'effet recherché}$$

Plus ce ratio est faible, plus la politique évaluée est considérée comme efficiente.

¹⁰¹ Chetty R., Friedman J., Hilger N., Saez E., Whitmore Schanzenbach D. et Yagan D. (2011), "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star", *The Quarterly Journal of Economics*, Vol. 126(4), p. 1593-1660.

¹⁰² Fredriksson P., Öckert B. et Oosterbeek H. (2013), "Long-Term Effects of Class Size", *The Quarterly Journal of Economics*, Vol. 128(1), p. 249-285.

Dans une étude consacrée aux effets du dispositif « zéro charges » Cahuc et al. (2017)¹⁰³ ont utilisé ce ratio pour évaluer l'efficacité du dispositif. La mesure, instaurée le 4 décembre 2008 pour une période limitée (initialement jusqu'au 31 décembre 2009, puis prolongée jusqu'au 30 juin 2010), était une aide à l'embauche destinée aux entreprises de moins de dix salariés. Ce dispositif a permis à ces entreprises d'être exonérées totalement de cotisations patronales pour les salariés nouvellement embauchés et payés au SMIC. Au-delà du SMIC, l'aide devenait dégressive pour s'annuler à partir d'une rémunération supérieure à 1,6 fois le montant du SMIC. La mise en place de ce dispositif n'a pu être prévue par les employeurs car elle est intervenue au moment même de son annonce publique. Ce critère lui confère un caractère de quasi-expérience.

Pour estimer les effets de ce dispositif sur l'évolution de l'emploi au cours de l'année 2009, Cahuc, Carcillo et Le Barbanchon (2017) ont utilisé une procédure de régression avec discontinuité au seuil de dix salariés. Plus précisément, ils ont comparé cette évolution dans les entreprises dont l'effectif était compris entre 6 et 10 salariés (le groupe cible), et dans celles dont l'effectif était compris entre 10 et 14 salariés (le groupe témoin). Ces auteurs ont montré que le dispositif avait permis de faire croître l'emploi de 0,8 % dans les entreprises de moins de 10 salariés. Cette hausse est certes quantitativement faible, mais elle est significative au sens statistique du terme. Selon les calculs des auteurs, le coût net d'un emploi créé a été quasiment nul, ce qui plaide bien évidemment en faveur de l'efficacité à court terme du dispositif « zéro charges ». Pour établir ce résultat, Cahuc et al. (2017) ont tout d'abord calculé le coût par emploi créé grâce au dispositif : ce coût est égal à l'opposé du rapport entre le salaire et l'élasticité de l'emploi à la variation du coût moyen de l'emploi résultant du crédit à l'embauche. Selon cette formule, le coût mensuel d'un emploi créé grâce au dispositif est estimé à 700 euros. Ce coût est un coût brut, car il ne tient pas compte des dépenses d'assurance-chômage qu'il aurait fallu verser aux travailleurs embauchés grâce à cette mesure et qui, sans cela, seraient restés au chômage. Grâce à une enquête conduite par Pôle Emploi à la fin de l'année 2009, Cahuc *et al.* (2017) ont estimé que ces dépenses évitées étaient du même montant que le coût brut de l'emploi créé grâce à la mesure. Au total, le coût net par emploi créé est quasiment nul.

Le laboratoire J-Pal a décidé de privilégier l'évaluation coûts-efficacité plutôt que les analyses coûts-bénéfices pour comparer l'efficacité de politiques évaluées dans le cadre de programmes expérimentaux et dont les résultats peuvent diverger¹⁰⁴.

Dans le cas des politiques de santé, l'efficacité peut être mesurée par le nombre de vies sauvées grâce à la nouvelle thérapie au cours de l'année ou des années qui suivent son introduction. Cet effet est mesuré en différentiel par rapport à la pratique courante. Une autre mesure de l'efficacité est l'augmentation de l'espérance de vie permise par la mise

¹⁰³ Cahuc P., Carcillo S. et Le Barbanchon T. (2017), "The Effectiveness of Hiring Credits", CEPR Discussion Paper no. 12537, à paraître dans *The Review of Economic Studies*. Pierre Cahuc a présenté cette étude devant le groupe de travail.

¹⁰⁴ Dhaliwal. I., Duflo E., Glennerster R. et Tulloch C. (2013), "Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries: A General Framework with Applications for Education", dans *Education Policy in Developing Countries*, University of Chicago Press, p. 285-338.

en œuvre de la nouvelle thérapie. Cependant, ces mesures ne tiennent pas compte de la qualité de vie du patient au cours de ces années gagnées. C'est pour cette raison que les économistes de la santé préfèrent utiliser les indicateurs Qaly ou Daly¹⁰⁵ qui tiennent compte simultanément de l'augmentation de l'espérance de vie et de la qualité de vie ou de l'absence d'incapacité du patient.

L'évaluation de l'indice Qaly est effectuée soit par des questionnaires administrés auprès des patients, soit par l'utilisation des valeurs obtenues par des questionnaires comme l'EQ-5D (EuroQol-5 Dimensions)¹⁰⁶, qui retient cinq dimensions pour mesurer les effets du traitement : la mobilité et l'autonomie des personnes, leurs difficultés dans les activités courantes, la douleur et l'anxiété.

Le Qaly fournit ainsi une mesure de l'utilité d'une année de vie supplémentaire en bonne santé. Il est par ailleurs utilisé pour calculer la valeur monétaire d'une intervention ou d'un traitement thérapeutique. Une année en bonne santé correspond à un Qaly de 1 ; une intervention causant la mort correspond à un Qaly de 0 ; une année au cours de laquelle l'intervention thérapeutique permet de prolonger la vie mais au prix d'une dégradation des conditions de vie (par exemple, en évitant le décès au prix d'un handicap) sera comptée entre 0 et 1, à un niveau fonction de l'ampleur de cette dégradation.

Le Qaly peut être utilisé pour comparer des traitements thérapeutiques concurrents, ces traitements pouvant avoir des effets différents à la fois sur l'espérance de vie et sur la qualité de vie des patients. Son intérêt ne réside donc pas seulement dans l'évaluation isolée d'un traitement thérapeutique, mais dans la possibilité de comparer les stratégies entre elles.

Que retenir ?

Lorsque la VAN repose sur des hypothèses et des estimations trop incertaines, et/ou lorsque l'intervention évaluée a un objectif principal à court terme et des effets attendus à long terme qui sont faibles il est préférable d'utiliser le ratio coûts-efficacité, défini comme le coût par unité d'effet obtenu.

Plus ce ratio est faible, plus la politique évaluée est considérée comme efficiente.

Dans le cas des politiques de santé, l'effet souvent utilisé est le nombre de vies sauvées à l'aide de l'intervention au cours de l'année ou des années suivantes.

L'indicateur Qaly permet de tenir compte simultanément de la qualité de vie du patient, en plus des gains d'espérance de vie, et des coûts du traitement.

¹⁰⁵ Qaly pour *Quality-Adjusted Life Year*, que l'on peut traduire par « année de vie pondérée par la qualité, » Daly pour *Disability-adjusted Life year*, que l'on peut traduire par année de vie corrigée de l'incapacité.

¹⁰⁶ Le questionnaire EQ-5D a été introduit en 1990 par le groupe EuroQol qui a été fondé en 1987 par des chercheurs de plusieurs disciplines et originaires de cinq pays européens : Pays-Bas, Royaume-Uni, Suède, Finlande et Norvège.

6. Les variables de contexte et le scénario de référence

Parmi ces variables, citons la situation macroéconomique (croissance, chômage, etc.), la démographie, le vieillissement de la population, les modifications de la structure de la population active, des niveaux de diplôme, la variation de l'âge de départ à la retraite, etc. Toutes ces variables ne doivent pas être nécessairement prises en compte simultanément. L'analyste pourra les distinguer séparément dans le cadre du chiffrage d'un scénario particulier.

Cette dernière remarque nous conduit à souligner qu'un calcul socioéconomique, du type de l'analyse coûts-bénéfices, ne doit pas déboucher sur un chiffrage unique, même lorsqu'il s'agit de l'appliquer à une politique particulière et non à plusieurs alternatives. Notre conseil est de livrer un tableau récapitulatif pour différents scénarios dépendant des valeurs retenues pour les variables de contexte. On trouvera des exemples de ce type de tableau dans l'article de Krueger et Whitmore (2001) sur la réduction de la taille des classes qui est présenté dans la sous-section 4.1 (ci-après), ou dans l'article d'Avery et Turner (2012)¹⁰⁷ sur le remboursement des prêts aux étudiants en fonction de leurs salaires futurs et de leurs filières de formation.

Le calcul socioéconomique ne s'utilise pas seulement de manière prospective, il peut l'être aussi rétrospectivement. Pour autant que l'on dispose des données longitudinales nécessaires (et ceci est plus facile sur les vingt ou trente dernières années, comme dans les DADS ou les enquêtes sur l'emploi de l'Insee) et d'une modélisation adéquate, on peut estimer ce qu'aurait produit une variation des paramètres d'une politique (par exemple, une modification des seuils donnant accès aux allègements de charges salariales), ou encore les effets d'une politique alternative (par exemple, le maintien de la dégressivité des allocations de chômage dans le cadre de la réforme de 2002).

Que retenir ?

Les variables de contexte incluent la situation macroéconomique, le taux de croissance de l'économie, l'évolution prévue du taux de chômage, la démographie, le vieillissement de la population, les modifications de la structure de la population active, la variation de l'âge de départ à la retraite, etc.

Toutes ces variables ne doivent pas être nécessairement prises en compte simultanément. L'analyste pourra les distinguer séparément dans le cadre du chiffrage d'un scénario particulier.

En raison des hypothèses multiples sur les variables de contexte, l'analyse coûts-bénéfices ne doit pas déboucher sur un chiffrage unique.

¹⁰⁷ Avery C. et Turner S. (2012), "Student Loans: Do College Students Borrow Too Much--Or Not Enough?", *Journal of Economic Perspectives*, Vol. 26(1), p. 165-192.

7. Les effets induits

Le problème de la prise en compte des effets induits soulève des questions similaires à celles relatives à la prise en compte des effets liés, évoquée précédemment au 3.III.2. Parmi les effets induits des politiques sociales ou éducatives, les plus souvent cités sont ceux qui concernent des dimensions connexes à la réussite scolaire, à l'emploi et aux revenus d'activité, telles que l'état de santé, la propension à la délinquance, ou encore le recours aux aides publiques et à l'assistance sociale. Il est important d'établir, par le biais des études statistiques ou sociologiques, la liste complète de ces effets induits, même s'ils sont difficiles à quantifier et encore plus à monétiser. Les recherches futures devront examiner de plus en plus ces questions, comme certaines le font d'ores et déjà. Par exemple, il est raisonnable de penser qu'une politique qui accroît les compétences cognitives et non cognitives des enfants peut avoir des effets sur la santé, la propension à la délinquance, le comportement citoyen, etc., des bénéficiaires de cette politique.

Les évidences empiriques sont les suivantes :

- les interventions expérimentales en faveur de la petite enfance défavorisée (le Perry Preschool Program, l'Abecedarian / Care Program et le programme Chicago Child Parent Center, tous les trois mis en œuvre aux États-Unis¹⁰⁸) et, de manière générale, les politiques éducatives qui ont eu pour objectif d'accroître le niveau général d'éducation de la population¹⁰⁹ diminuent le taux de délinquance à l'âge adulte (notamment entre 18 et 25 ans) ; pour donner un ordre de grandeur, la participation au programme expérimental Chicago Child Parent Center a réduit de 8 points de pourcentage la probabilité d'être arrêté à la suite d'un délit à l'âge de 18 ans ;
- les interventions précoces telles que le Perry Preschool Program ou l'Abecedarian / Care Program améliorent l'état de santé et les comportements en matière de santé des bénéficiaires à l'âge adulte¹¹⁰ ; on observe par ailleurs une forte corrélation (qui peut être interprétée comme une relation causale à l'aide de méthodes quasi expérimentales) entre l'augmentation de l'âge minimal de sortie du système scolaire et l'état de santé, ainsi que la mortalité, à l'âge adulte¹¹¹ ; dans ces dernières, une année supplémentaire

¹⁰⁸ Sur les effets de long terme du Perry Preschool Program, voir Heckman J., Moon S. H., Pinto R., Savelyev P. et Yavitz A. (2010), "The rate of return to the High Scope Perry Preschool Program," *Journal of Public Economics*, Vol. 94(1-2), p. 114-128 ; et Belfield C. R., Nores M., Barnett W. S. et Schweinhart L. (2006), "The High/Scope Perry Preschool program: Cost-benefit analysis using data from the age-40 follow-up", *Journal of Human Resources*, Vol. 41 (1), p. 162-190. Sur les effets de long terme du programme Abecedarian/Care, voir García J. L., Heckman J. J., Leaf D. E. et Prados M. J. (2017), "Quantifying the Life-cycle Benefits of a Prototypical Early Childhood Program", NBER Working Paper No. 23479. Sur les effets de long terme du Chicago Child Parent Center, voir Reynolds A. J., Temple J. A., White B. A., Ou S. R. et Robertson D. L. (2011), "Age-26 Cost-Benefit Analysis of the Child-Parent Center Early Education Program", *Child Development*, Vol. 82(1), p. 379-404.

¹⁰⁹ L. Lochner, (2004), "Education, work and crime: A human capital approach", *International Economic Review*, Vol. 45(3), p. 811-843; L. Lochner et E. Moretti (2004), "The effect of education on crime: Evidence from prison inmates, arrests, and self-report", *American Economic Review*, Vol. 94(1), p. 155-189; B. Bell, R. Costa et S. Machin (2018), "Why does education reduce crime?", IZA Discussion Paper No. 11805, Bonn.

¹¹⁰ Conti G., Heckman J. et Pinto R. (2016), "The effects of two influential early childhood interventions on health and healthy behavior", *The Economic Journal*, Vol. 126(596), p. 28-65.

¹¹¹ Albouy V. et Lequien L. (2009), "Does compulsory education lower mortality?", *Journal of Health Economics*, Vol. 28(1), p. 155-168 ; Clark D. et Royer H. (2013), "The Effect of education on adult mortality and health: Evidence from Britain", *American Economic Review*, Vol. 103(6), p. 2087-2120.

d'éducation est associée (de manière causale) à une amélioration de l'état de santé auto-déclaré de 5 à 10 points sur une échelle allant de 0 à 100 ;

- les bénéficiaires du programme Perry Preschool étaient à l'âge de 40 ans moins nombreux à avoir bénéficié de l'aide sociale au cours des dix années précédentes ;
- un niveau d'éducation plus élevé accroît le civisme, notamment la participation aux élections et l'engagement citoyen ; au-delà de la corrélation souvent observée, certaines études quasi expérimentales concluent à un effet causal de l'éducation sur le civisme¹¹².

Lorsque l'on considère les effets induits d'une politique d'investissement social, il faut prêter attention au problème des doubles comptes. Ce problème résulte du fait que certaines variables de résultats reflètent en réalité le même phénomène, mais mesuré différemment : par exemple, une amélioration des résultats scolaires en fin de second cycle et une plus grande probabilité d'avoir un diplôme de l'enseignement supérieur peuvent être considérés comme la conséquence d'une même cause, à savoir une augmentation des capacités cognitives induites par une réduction de la taille des classes à l'école élémentaire. Si ces effets sont simultanément monétisés dans le calcul des bénéfices et que l'on additionne les euros gagnés grâce à de meilleures notes et ceux reflétant la valeur d'obtenir ce diplôme, l'estimation risque de surestimer, ou tout au moins de mésestimer, le véritable rendement social du programme. Dans le chapitre suivant, nous évoquerons la façon dont le Washington State Institute for Public Policy propose de traiter ce problème.

Les analyses coûts-bénéfices (ACB) ont certes l'avantage de rendre compte de certains éléments importants tels que le temps requis pour recouvrer les dépenses engagées ou bien l'identification des personnes pour lesquelles le programme a des effets particulièrement bénéfiques. Mais là n'est pas l'avantage principal des ACB : celles-ci permettent de valoriser les effets à moyen ou long terme en matière de santé, de recours à l'aide sociale, de délinquance ou d'éducation. Pour le dire autrement, les effets induits, notamment non marchands, d'un programme peuvent être monétisés et additionnés de façon à pouvoir comparer la somme des bénéfices identifiés aux coûts engagés.

Les interventions préscolaires auprès de la petite enfance sont les interventions pour lesquelles nous disposons le plus souvent de l'observation d'effets à long terme. Les travaux réalisés notamment sur les programmes Perry Preschool et Abecedarian livrent des conclusions importantes sur la portée sociale et l'analyse coûts-bénéfices de telles interventions. En économie, les études souvent citées que James Heckman et ses collaborateurs ont consacrées à ces programmes, en particulier au programme Perry Preschool, fournissent des exemples de la façon dont les effets non marchands de ces interventions peuvent être monétisés.¹¹³ L'étude d'Heckman *et al.* (2010) relative au programme Perry Preschool est présentée en Annexe A.V.

¹¹² Milligan K., Moretti E. et Oreopoulos P. (2004), "Does education improve citizenship? Evidence from the United States and the United Kingdom", *Journal of Public Economics*, Vol. 88(9-10), p. 1667-1695 ; Siedler T. (2010), "Schooling and citizenship in a young democracy: Evidence from postwar Germany", *Scandinavian Journal of Economics*, Vol. 112(2), p. 315-338.

¹¹³ Heckman J., Moon S. H., Pinto R., Savelyev P. et Yavitz A. (2010), "The rate of return to the High Scope Perry Preschool Program", *Journal of Public Economics*, Vol. 94(1-2), p. 114-128 ; García J. L., Heckman J. J., Leaf D. E., et Prados M. J. (2017), "Quantifying the Life-cycle Benefits of a Prototypical Early Childhood Program", NBER

Dalziel *et al.* (2015)¹¹⁴ ont publié une synthèse des analyses coûts-bénéfices et des effets de long terme des programmes préscolaires destinés en particulier à la petite enfance la plus exposée aux difficultés sociales ou d'apprentissage. Après avoir identifié, à partir de toutes les bases de données disponibles en 2013¹¹⁵, les 147 études ayant proposé à cette date des analyses coûts-bénéfices des programmes de ce type, ils retiennent pour leur méta-analyse seulement treize évaluations socioéconomiques répondant à des critères stricts¹¹⁶, à savoir :

- elles devaient reposer à la fois sur l'estimation des coûts et des bénéfices attendus ;
- elles devaient permettre d'estimer ces éléments conjointement pour un groupe cible mais aussi pour un groupe de contrôle ne bénéficiant pas de l'intervention ;
- les interventions évaluées devaient être réalisées dans des centres dédiés ;
- elles devaient concerner des enfants de moins de 5 ans vivant dans des milieux sociaux défavorisés.

Ces treize études concernent au total six programmes différents : le High-Scope Perry Preschool, mis en œuvre de 1962 à 1967 à Ypsilanti (Michigan) ; le Carolina Abecedarian, mis en place entre 1972 et 1977 à Chapel Hill (Caroline du Nord) ; le Chicago Child-Parent Center Program (CCPC ci-après), mis en œuvre entre 1983 et 1985 à Chicago (Illinois) ; l'Even Start Family Literacy Program, généralisé aux États-Unis à partir de 1989 ; l'Early Head Start, généralisé aux États-Unis à partir de 1995 ; le Sure Start, généralisé en Angleterre entre 1999 et 2003.

Seuls les trois premiers de ces programmes – qui sont aussi les plus anciens – ont produit des bénéfices très supérieurs à leurs coûts. Selon Dalziel *et al.* (2015), la cause première de cette divergence est la qualité variable des programmes examinés : ceux dont le contenu était plus intense et mieux défini, en général réservés à des groupes d'enfants présentant des difficultés spécifiques d'apprentissage, ont produit les meilleurs résultats. **Les programmes mis en place à plus grande échelle, en général moins ciblés, tels que l'Early Head Start ou le Sure Start, n'ont pas eu les mêmes rendements sociaux**¹¹⁷. Les évaluations divergentes peuvent également provenir de périodes d'observation plus ou

Working Paper No. 23479 ; Heckman J., Humphries J. et Veramendi G. (2017), "The Non-Market Benefits of Education and Ability", NBER Working Paper No 23896.

¹¹⁴ Dalziel K. M., Halliday D. et Segal L. (2015), "Assessment of the Cost-Benefit Literature on Early Childhood Education for Vulnerable Children: What the Findings Mean for Policy", *Sage Open*, Vol. 5(1).

¹¹⁵ Les auteurs de l'étude ont plus particulièrement exploré les bases Medline, Embase, A+Education, Family Databases, Proquest Education journals, Proquest Psychology journals, Education Research Complete, ERIC, EconLit, PsychINFO et PsychArticles, en utilisant les mots clefs « child care, day care, kindergarten, preschool, nursery, early intervention education, benefit-cost ».

¹¹⁶ Indépendamment de leur application aux politiques en faveur de la petite enfance, ces critères peuvent sembler trop stricts (notamment le troisième). Ceux du Washington State Institute for Public Policy (WSIPP, cf. Chapitre 4) le sont moins, mais le WSIPP ne retient dans ses méta-analyses que les études nord-américaines. Il y a toutefois une intersection assez large entre ces différentes approches : un consensus se dégage autour de la nécessité de retenir seulement des ACB reposant sur des expérimentations contrôlées ou des quasi-expérimentations rigoureusement effectuées. Cela dit, de plus en plus de chercheurs plaident pour la vérifiabilité des résultats de ces expériences ou quasi-expériences et leur reproductibilité (cf. Chapitre 2).

¹¹⁷ En termes de ciblage des politiques, on voit ici apparaître une analogie avec certaines des conclusions de la méta-analyse de Card, Kluve et Weber (2018) citée plus haut.

moins longues, les effets de long terme apparaissant plus naturellement et étant mieux appréhendés lorsque le programme donne lieu à un suivi longitudinal plus étendu.

Dalziel *et al.* (2015) utilisent les deux mesures habituelles d'efficience, la valeur actualisée nette et le ratio bénéfices/coûts (parfois ci-après dénommé « rendement »). Les coûts graduels (incrémentaux) du programme incluent les salaires des éducateurs et des autres personnels requis pour la mise en œuvre du programme, les équipements des centres dédiés (incluant le mobilier des écoles, des salles d'accueil, etc.), les matériels pédagogiques (jeux, livres, guides, etc.), ainsi que les autres composantes entrant dans les coûts des programmes. Les bénéfices supplémentaires apportés par le programme incluent le moindre recours aux cours de rattrapage et à l'éducation spécialisée, l'augmentation de la participation des parents au marché du travail, l'accès plus fréquent des enfants à l'éducation secondaire supérieure (*i.e.*, au lycée), le moindre recours à l'aide sociale des bénéficiaires arrivés à l'âge adulte en lien avec leur plus grande employabilité et leur propension plus faible à la délinquance. Une fois l'ensemble de ces coûts et bénéfices monétisés (mesurés ici en dollars US de 2011), il est possible de calculer le ratio entre les sommes actualisées des bénéfices et des coûts, la valeur actualisée nette définie comme la différence entre ces deux quantités ou le taux de rendement interne défini comme le taux qui égalise les valeurs présentes actualisées des flux de bénéfices et de coûts.

Peu d'études consacrées aux six programmes ci-dessus listés contiennent des analyses de sensibilité. Quand elles le sont, c'est principalement en faisant varier le taux d'actualisation. Seules deux études présentent des analyses de sensibilité plus poussées, notamment en tenant compte des coûts de financement de ces programmes par le biais de l'impôt.¹¹⁸ Il s'agit de l'article de Reynolds *et al.* (2011) relatif au programme CCPC et celui d'Heckman *et al.* (2010) consacré au programme High-Scope Perry Preschool¹¹⁹.

Les coûts par enfant varient considérablement d'un programme à l'autre, en raison de leur intensité et de leurs contenus très différents : ils allaient de 5 980 US dollars pour l'Even Start (7 870 US dollars pour le Sure Start) jusqu'à 45 190 US dollars pour le programme Abecedarian (29 836 US dollars pour le *Perry Preschool*). Les ratios bénéfices/coûts ne sont pas pour autant favorables aux programmes les moins coûteux. Les programmes Perry Preschool et CCPC ont été évalués dans plusieurs études, qui ont conclu à des efficacités différentes. Par exemple, selon Belfield *et al.* (2006),¹²⁰ le Perry Preschool aurait eu un rendement très élevé, égal à 16,1,¹²¹ alors que pour Heckman *et al.* (2010), son rendement aurait été plus faible, égal à 6,20. Le rendement du programme Abecedarian aurait été compris entre 3,78 et 10,15, celui du programme CCPC entre 4,82 et 10,83. Les trois autres

¹¹⁸ Ces coûts correspondent à ce que les économistes appellent les « *deadweight losses* », terme que l'on peut traduire par « pertes sèches » ou « pertes d'efficacité ».

¹¹⁹ Reynolds A. J., Temple J. A., White B. A., Ou S. R. et Robertson D. L. (2011), "Age-26 Cost-Benefit Analysis of the Child-Parent Center Early Education Program", *Child Development*, Vol. 82(1), p. 379-404 ; Heckman J., Moon S. H., Pinto R., Savelyev P. et Yavitz A. (2010), "The rate of return to the High Scope Perry Preschool Program," *Journal of Public Economics*, Vol. 94(1-2), p. 114-128.

¹²⁰ Belfield C. R., Nores M., Barnett W. S. et Schweinhart L. (2006), "The High/Scope Perry Preschool program: Cost-benefit analysis using data from the age-40 follow-up", *Journal of Human Resources*, Vol. 41 (1), p. 162-190.

¹²¹ Ce qui signifie qu'il aurait rapporté 16,1 dollars par dollar investi.

programmes recensés, Even Start, Early Head Start et Sure Start, auraient eu des rendements proches de zéro ; ils étaient pourtant les moins coûteux.

Outre ces mesures d'efficience, l'intérêt de la méta-analyse réalisée par Dalziel *et al.* (2015) est de montrer que le rendement d'une intervention est d'autant plus élevé que sa mesure tient compte des effets de long terme, notamment en monétisant les effets induits non marchands. Cela suppose toutefois que ces interventions aient donné lieu à des suivis sur de longues périodes. Ainsi, les analyses coûts-bénéfices du programme Perry Preschool cités préalablement ont montré que son rendement dépend très fortement de la prise en compte de la réduction des coûts résultant d'une moindre délinquance à l'âge adulte. Les ACB de l'Abecedarian n'ont malheureusement pas pu confirmer ce résultat, car le suivi de ce programme ne permettait pas d'observer les délits commis par les participants à l'expérience. À l'inverse, ce dernier programme était le seul dont le suivi fournissait des informations sur la dépendance tabagique, et permettait donc d'évaluer les risques de ce comportement sur la santé. L'ACB du programme CCPC réalisée par Reynolds *et al.* (2011)¹²² est celle qui repose sur le plus grand nombre d'effets induits : leur analyse intègre en effet plusieurs mesures de réussite scolaire, le nombre de délits ayant donné lieu à des suites judiciaires, des mesures de la toxicomanie et de la santé mentale à 21 ans, etc. C'est pour cette raison que ce rapport recommande dès le début de l'évaluation de lister tous les effets possibles, de les quantifier si possible et de valoriser (donner une valeur monétaire ou monétariser) si possible aux effets quantifiés. La VAN socioéconomique inclut seulement les effets quantifiés et monétarisés.

Que retenir ?

Parmi les effets induits des investissements sociaux, les plus souvent cités sont ceux qui concernent des dimensions connexes à la réussite scolaire, à l'emploi et aux revenus d'activité, telles que l'état de santé, la propension à la délinquance, ou encore le recours aux aides publiques et à l'assistance sociale.

Les effets induits peuvent être monétisés et additionnés de sorte à pouvoir comparer la somme des bénéfices identifiés aux coûts engagés.

Lorsque l'on considère les effets induits d'une politique d'investissement social, il faut prêter attention au problème des doubles comptes. Ce problème vient du fait que des variables de résultats différentes mesurent en réalité un même phénomène.

Généralement, la VAN augmente avec le nombre d'effets induits pris en compte pour autant que ceux-ci soient correctement monétisés et ne reflètent pas le même phénomène.

C'est pourquoi il est recommandé de lister tous les effets possibles dès le début de l'évaluation.

¹²² Reynolds A. J., Temple J. A., White B. A., Ou S. R. et Robertson D. L. (2011), "Age-26 Cost-Benefit Analysis of the Child-Parent Center Early Education Program", *Child Development*, Vol. 82(1), p. 379-404.

8. Situation contrefactuelle et expérimentation

Le *Guide de l'évaluation socioéconomique des investissements publics*¹²³ de France Stratégie et la Direction générale du Trésor propose de décomposer la notion de « situation de référence » en deux parties : le « scénario de référence », qui renvoie au cadrage général indépendant de la réalisation de la politique publique considérée, et « l'option de référence », ou *situation contrefactuelle*, qui correspond à la situation qui prévaudrait en l'absence de mise en œuvre de cette politique. Cette distinction permet d'insister sur le fait que la situation contrefactuelle n'est pas nécessairement la même que la situation de référence ou, dit autrement, que l'alternative à la politique n'est pas toujours « ne rien faire ». Nous retenons également cette définition.

Le *traitement*, terme utilisé par analogie avec la bioclinique, est la situation engendrée par la mise en application de la politique sociale considérée, politique dont nous souhaitons mesurer les effets. Leur estimation est rendue difficile par le fait que **nous ne pouvons observer la situation contrefactuelle pour ceux des individus qui bénéficient de cette politique, de même que nous ne pouvons observer le traitement pour ceux qui n'en bénéficient pas**. L'objectif est alors de reconstituer des statistiques permettant de retrouver ces données manquantes ou, du moins, des valeurs qui s'en approchent.

Par le passé, nombre des évaluations conduites par les chercheurs et les administrations ne tenaient pas réellement compte de cette difficulté. L'approche longtemps retenue consistait à exploiter des données non expérimentales, provenant généralement d'enquêtes en population générale. Le principe était alors de construire un groupe témoin composé d'individus semblables du point de vue de leurs caractéristiques socioéconomiques observables aux individus du groupe cible, et d'estimer la différence moyenne de salaire, d'emploi, de santé, etc., dans ces deux groupes après mise en œuvre du programme. En réalité, cette différence est en fait la combinaison de deux effets, l'effet spécifique du programme que l'on cherche à mesurer, mais aussi le fait que les deux groupes d'individus peuvent se comporter différemment face au programme considéré. En effet, ce programme peut ne pas intéresser de la même façon tous les individus éligibles. Certains peuvent demander avec force à bénéficier du programme et des avantages qu'il procure, d'autres au contraire peuvent être réticents et ne pas souhaiter participer. En ce cas, la différence de résultats entre les deux groupes, cible et témoin, reflète en partie ces **comportements d'auto-sélection** : elle est entachée d'un biais d'estimation car, malgré leur similarité apparente, les groupes cibles et témoin diffèrent sur un point comportemental central pour l'analyse. C'est pourquoi ce type de comparaison est de moins en moins admis pour les évaluations d'impact.

8.1. L'expérimentation aléatoire : avantages et limites

L'évaluation par expérimentation randomisée essaie de s'affranchir de ce type de biais en proposant d'affecter de manière aléatoire (par tirage au sort) les individus éligibles qui font

¹²³ Cf. *Guide de l'évaluation socioéconomique des investissements publics*, rédigé sous l'autorité du comité d'experts des méthodes d'évaluation socioéconomique des investissements publics présidé par Roger Guesnerie, décembre 2017, Direction générale du Trésor et France Stratégie.

partie de l'échantillon d'étude à un groupe de traitement qui bénéficiera du programme devant être évalué, ou bien à un groupe de contrôle qui n'en bénéficiera pas. La différence de résultats constatée après mise en œuvre du programme est supposée refléter fidèlement l'effet moyen du programme. En effet, en vertu de l'affectation aléatoire, le fait de pouvoir bénéficier ou non de la politique évaluée est indépendant de toutes choses autres que du hasard, notamment des préférences, du niveau d'éducation, des caractéristiques sociodémographiques, de la motivation, de l'aversion au risque, etc., et en particulier des déterminants susceptibles d'affecter à la fois les variables de résultat et l'accès à la politique. Dans ces conditions, la simple comparaison des groupes aléatoirement composés mesure l'impact causal de la politique étudiée puisque la seule différence *a posteriori* entre les groupes ne peut être que la conséquence de cette politique sur les comportements, en vertu de la procédure d'affectation aléatoire. Cette méthode permet de mesurer des effets hétérogènes et de recouvrer plusieurs statistiques d'intérêt, tel l'effet du traitement sur les traités.

Cependant, bien qu'elle soit jugée comme préférable aux méthodes d'évaluation d'impact non ou quasi expérimentales¹²⁴, l'expérimentation contrôlée n'est pas pour autant une méthode d'évaluation parfaite. Elle peut elle-même être sujette à certains types de biais, au premier rang desquels figurent les biais d'adhésion (*non compliance*) et de substitution (voir *infra*). Ces biais ont pour conséquence que le groupe cible (ou groupe de traitement) et le groupe témoin (ou groupe de contrôle) diffèrent sensiblement au terme de l'expérience.

Le biais d'adhésion apparaît lorsque des individus éligibles, choisis aléatoirement pour participer au groupe cible, refusent d'y participer ou bien acceptent puis cessent de participer à l'expérience avant que celle-ci parvienne à son terme. Ainsi, il est à craindre que les individus qui refusent de participer à l'évaluation expérimentale d'une nouvelle politique soient précisément ceux qui pensent en tirer le moins d'avantages. En ce cas, la différence des moyennes de résultat entre les groupes de traitement et de contrôle ne mesure pas vraiment ce que l'on souhaite mesurer, à savoir l'effet moyen de la politique, mais l'effet moyen de cette politique pour les individus qui ont accepté de participer à l'expérience. L'écart entre ces deux quantités peut être assez substantiel, comme l'ont montré des expériences conduites dans différents domaines¹²⁵.

En présence d'un biais d'adhésion, deux solutions peuvent être envisagées. En premier lieu, il est toujours possible de comparer les résultats des groupes en fonction de leur affectation initiale, et ce indépendamment de leur situation réelle vis-à-vis de l'intervention. Cette comparaison identifie l'effet moyen de ce que l'on appelle **l'intention de traitement (*intention to treat*)**. Si le protocole a vraiment été mis en place de façon

¹²⁴ Les méthodes quasi expérimentales les plus fréquentes sont la régression avec discontinuité (*regression discontinuity design*), la méthode des doubles différences (*differences-in-differences*) et la technique d'appariement d'échantillons (*matching*). Ces méthodes ont été décrites de façon non technique par R. Desplat et M. Ferracci dans le document « Comment évaluer l'impact des politiques publiques ? Un guide à l'usage des décideurs et praticiens » (France Stratégie, septembre 2016).

¹²⁵ Parmi les expérimentations faisant apparaître cet écart, celle-ci est souvent citée : Heckman J., Hohmann N., Smith J. et Khoo M. (2000), « Substitution and Dropout Bias in Social Experiments: A Study of an Influential Social Experiment », *The Quarterly Journal of Economics*, Vol. 115(2), p. 651-694.

aléatoire, ce résultat ne souffre pas de biais : il mesure l'effet de la possibilité d'accéder à la politique. Dans certains cas, cette quantité est intéressante en soi pour le décideur puisque, en dehors d'un dispositif expérimental, les politiques publiques sont souvent proposées sans obligation d'adhésion. Néanmoins, cette quantité ne permet pas à elle seule d'identifier l'effet de la politique sur les individus effectivement traités. Il reste possible d'identifier l'effet moyen du traitement pour ceux qui adhèrent pleinement au moyen d'hypothèses supplémentaires. La première hypothèse stipule qu'une partie significative des individus souscrit au traitement lorsqu'il leur est proposé. La seconde signifie qu'aucun individu n'irait systématiquement à l'encontre du traitement qui lui est proposé. Sous ces deux hypothèses, l'effet moyen du traitement sur ceux qui y adhèrent est obtenu en divisant la différence de résultat moyen entre les groupes (traitement et contrôle) par la différence des probabilités d'adhésion dans ces deux groupes. L'effet moyen du traitement et l'effet moyen sur les traités peuvent être obtenus à partir d'hypothèses structurelles¹²⁶ ou en faisant l'hypothèse que l'effet du traitement est le même pour tous.

Le biais de substitution peut apparaître dès lors que certains membres du groupe de contrôle, n'ayant pas accès au programme durant le temps de l'expérience, peuvent accéder à des substituts au programme avant la fin de l'expérience. Pensons à un programme éducatif qui accroît le nombre d'heures de mathématiques. Il est possible que les parents des élèves des classes faisant partie du groupe de contrôle recourent à des cours privés. Dès lors, l'évaluation expérimentale de la politique évaluée comparera la situation après intervention à un contrefactuel qui n'est pas nécessairement celui pertinent pour le décideur. Dans l'exemple ci-dessus, on peut penser que le contrefactuel pertinent est un nombre d'heures moins élevé de mathématiques.

Ce problème est à distinguer des problèmes liés à la mise en œuvre d'une expérimentation¹²⁷, tels les effets Hawthorne et Henry. Le premier type d'effet caractérise la réaction particulière des personnes participant à l'expérience qui, se sachant observées par des analystes (statisticiens, biologistes, économètres, etc.) et/ou des politiciens dans le cadre d'une expérimentation, ont une attitude différente de celle qu'ils auraient dans un cadre non expérimental. Le second type d'effet désigne l'éventuelle réaction du groupe de contrôle qui, ne pouvant pas profiter d'une politique potentiellement bénéfique, ne se comporte pas comme s'il était dans la situation où cette politique n'est pas envisagée. On peut ainsi imaginer dans notre exemple que les professeurs en charge des classes, hostiles au principe de l'expérimentation, décident unilatéralement d'intensifier le contenu de l'enseignement de mathématiques. Il est clair qu'il est difficile de leur interdire d'agir ainsi, leur argument étant de ne pas défavoriser les élèves qui n'ont pas accès à un programme qu'ils supposent bénéfique. Là encore, réduire ce biais, qui tend à sous-estimer les effets du programme ou de la politique éducative que l'on souhaite évaluer, est chose difficile. En conséquence, la seule solution pragmatique, quoique difficile à mettre en œuvre, est de limiter les possibilités d'accès à des substituts au sein du groupe de contrôle.

¹²⁶ À l'aide d'un modèle de Roy généralisé, par exemple.

¹²⁷ Ces problèmes ne sont certes pas spécifiques à la procédure de randomisation.

Au total, et en dépit de ces biais potentiels, l'identification des effets à court et plus long termes des politiques sociales et éducatives devrait, autant que faire se peut, provenir d'expérimentations randomisées, ou pour le moins de méthodes quasi expérimentales. L'expérimentation randomisée réconcilie action de terrain et recherche ; elle peut aider à combler l'incompréhension que les citoyens manifestent à l'endroit de la recherche en sciences sociales. Son principe est simple à comprendre, malgré les sentiments de frustration, voire d'injustice, qu'elle peut induire non seulement chez les personnes éligibles et non sélectionnées, mais aussi chez les personnels chargés de la mise en œuvre (salariés du service public de l'emploi, corps médical...). Elle a par ailleurs l'avantage de fournir des résultats dont la validité interne est plus grande que beaucoup d'autres méthodes d'évaluation.

8.2. L'analyse coûts-bénéfices d'un programme expérimental

Comment appliquer le calcul socioéconomique à une politique d'investissement social dont les effets sont observés à l'aide d'une expérimentation contrôlée ? L'article de Kline et Walters (2016)¹²⁸ consacré aux effets d'une expérimentation du programme Head Start (voir encadré) contient à ce jour l'une des meilleures réponses à cette question. Leur étude fournit par ailleurs des estimations et une analyse coûts-bénéfices tenant compte des deux types de biais venant d'être décrits, à savoir les biais d'adhésion et de substitution.

Mis en place dès 1965 aux États-Unis, le programme Head Start fournit une éducation et un suivi préscolaire aux enfants de familles américaines défavorisées. Ce programme est financé par des fonds publics alloués à des organisations publiques ou privées œuvrant à l'échelon local. En principe, seuls sont éligibles à ce programme les enfants de 3 ou 4 ans dont les parents perçoivent un revenu inférieur au seuil fédéral de pauvreté. Toutefois, ce critère n'est pas strict. Ainsi, les enfants de familles bénéficiant du programme Temporary Aid for Needy Families (TANF) peuvent également candidater à Head Start. Le programme Head Start, gratuit pour les familles, est très demandé, et ne peut offrir des places à tous les enfants éligibles. Ainsi, en 2002, seuls 85 % des enfants y ayant droit ont pu y participer.

Head Start n'est pas le seul programme préscolaire offert aux enfants de familles défavorisées. Par exemple, le Child Care Development Fund verse aux autorités locales des donations finançant des programmes de garde d'enfant pour les familles à faible revenu, et distribue des bons leur permettant d'accéder à des centres d'éducation préscolaire. Les fonds du programme TANF cité plus haut permettent plus ou moins la même chose. Dès lors que ces possibilités existent, il est clair que **le contrefactuel à Head Start n'est pas l'absence de préscolarisation, mais une situation mixte qui mêle garde à domicile et participation à des programmes substitués. Cela complique l'évaluation de Head Start et de manière générale de tout programme coexistant avec des substitués.**

¹²⁸ Kline P. et Walters C. (2016), "Evaluating Public Programs with Close Substitutes: The Case of Head Start", *The Quarterly Journal of Economics*, Vol. 131(4), p. 1795-1848.

Encadré – Le programme Head Start

« Head Start est un programme du Département de la Santé, de l'éducation et des services sociaux des États-Unis qui fournit une éducation complète, des services d'implication parentale, de santé, de nutrition, aux enfants à faibles revenus et à leurs familles. Head Start a été créé en 1965 et fut modifié par le Head Start Act de 1981. Ce programme a traversé les modifications les plus profondes dans son renouvellement de décembre 2007. C'est aussi le programme ayant la plus grande longévité parmi ceux destinés à régler la pauvreté systémique aux États-Unis. Depuis sa création et jusqu'en 2005, plus de 22 millions d'enfants en âge préscolaire ont participé au programme Head Start. En 2005, Le budget annuel de 6,8 milliards de dollars a permis d'offrir ces services à plus de 905 000 enfants, 57 % d'entre eux avaient 4 ans ou plus, et 43 % avaient 3 ans ou moins. Les services furent fournis par 1 604 programmes différents opérant dans plus de 48 000 classes réparties dans chaque État (et presque dans chaque comté) pour un coût moyen de 7 222 dollars par enfant. Le personnel est composé de presque 212 000 personnes salariées en complément des bénévoles qui sont six fois plus nombreux ».

Source Wikipédia

Plusieurs travaux académiques exploitant des données non expérimentales ont évalué les effets du programme Head Start en comparant les résultats scolaires à court et plus long termes des enfants en ayant bénéficié à ceux d'enfants comparables qui n'ont pas participé au programme. En général, les résultats de ces études montrent que le programme a des effets positifs à court terme sur la réussite scolaire, mais aussi des effets bénéfiques sur des variables de plus long terme, telles que les salaires, la santé ou la délinquance¹²⁹. À l'opposé de ces études, l'expérimentation randomisée Head Start Impact Study (HSIS), conduite en 2002 auprès de plus de 4 000 enfants, a montré que la participation au programme Head Start accroît les capacités cognitives des enfants à court terme mais que cet effet décline dès l'entrée à l'école primaire¹³⁰.

Kline et Walters (2016) ont réexaminé les résultats de cette expérimentation randomisée. Pour ce faire, ils ont tenu explicitement compte de l'existence des programmes préscolaires autres que Head Start, et ont considéré que les enfants pouvaient être dans l'une des trois situations suivantes : être gardé au domicile des parents ou d'un proche, participer au programme Head Start, ou bien participer à un programme préscolaire autre que Head Start (cet autre programme est un *substitut* à Head Start). Au total, dans le cadre de l'expérimentation HSIS, Kline et Walters (2016) ont distingué cinq catégories d'enfants :

- ceux qui, quel que soit le nombre de places offertes dans le programme Head Start, y participent (appelés les « *always takers* », ou « toujours preneurs ») ;

¹²⁹ Parmi les études non expérimentales les plus significatives, on peut citer celles de Garces E., Thomas D. et Currie J. (2002=, "Longer-Term Effects of Head Start", *American Economic Review*, vol. 92, p. 999-1012 ; Deming D. (2009), "Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start", *American Economic Journal: Applied Economics*, vol. 1, p. 111-34, juillet ; et Carneiro P. et Ginja R. (2014), "Long-Term Impacts of Compensatory Preschool on Health and Behavior: Evidence from Head Start", *American Economic Journal: Economic Policy*, vol. 6, p. 135-73, novembre.

¹³⁰ Voir Puma M. *et al.*, (2010), "Head Start Impact Study. Final Report", Washington DC: U.S. Department of Health and Services, Administration for Children and Families.

- ceux qui étaient précédemment dans un programme préscolaire autre que le Head Start mais qui décident de participer à Head Start lorsque le nombre de places offertes dans ce programme augmente (appelés les « *compliers* » préalablement participant à un programme substitut) ;
- ceux qui étaient précédemment gardés à domicile mais qui décident de participer à Head Start lorsque le nombre de places offertes dans ce programme augmente (appelés les « *compliers* » préalablement gardés à domicile) ;
- ceux qui étaient précédemment dans un programme préscolaire autre que le Head Start mais qui décident de ne pas participer à Head Start lorsque le nombre de places offertes dans ce programme augmente (appelés les « *non-compliers* » participant à un programme substitut) ;
- et ceux qui étaient précédemment gardés à domicile mais qui décident de ne pas participer à Head Start lorsque le nombre de places offertes dans ce programme augmente (appelés les « *non-compliers* » gardés à domicile).

En présence de programmes substitués et de non-adhésion de certaines familles éligibles (« *non compliance* »), l'estimateur expérimental habituel (soit la différence des résultats des enfants du groupe cible et de ceux des enfants du groupe témoin) ne peut plus être directement interprété comme l'effet moyen du programme Head Start. En ce cas, seul peut être interprété et estimé correctement l'effet moyen de Head Start sur les résultats scolaires de ceux des enfants qui participaient préalablement à un autre programme préscolaire ou étaient gardés à domicile, et qui, lorsque le nombre de places offertes dans le programme Head Start augmente, décident d'y participer. Pour ce sous-groupe d'enfants (les « *compliers* »), l'effet estimé, noté $LATE_h$ par Kline et Walters (2016), est supérieur à celui estimé simplement en comparant les résultats moyens des enfants du groupe cible et ceux des enfants du groupe témoin.

Pour faire l'analyse coûts-bénéfices d'un programme expérimental de ce type (i.e., avec biais potentiels de substitution et de non-adhésion), Kline et Walters (2016) proposent un modèle simplifié qui est présenté brièvement dans l'encadré sur l'évaluation d'un programme expérimental. Ce modèle fournit un cadre d'analyse simple qui peut être assez facilement reproduit dans d'autres contextes.

**Encadré – Un modèle simple d'analyse coûts-bénéfices
pour l'évaluation d'un programme expérimental**

Pour conduire l'analyse coûts-bénéfices d'un programme préscolaire évalué par une expérimentation randomisée, Kline et Walters (2016) proposent un modèle théorique simplifié. Le programme considéré est le programme nord-américain Head Start, un programme fédéral qui propose un suivi éducatif aux enfants de milieux défavorisés dès l'âge de 4 ans (voir plus haut). Ils notent B le revenu total disponible (après impôt), dont la formule est supposée être :

$$B = B_0 + (1 - \tau)pE(y_i) \quad (*)$$

Dans cette expression, $E(y_i)$ est le résultat moyen de l'élève aux tests scolaires, p représente le prix de marché d'une unité de capital humain mesurée par le résultat scolaire¹³¹, τ est le taux d'impôt auquel fait face un enfant de famille éligible au programme Head Start lorsqu'il sera devenu adulte, et B_0 est une constante de normalisation (qui peut être interprétée comme le revenu total disponible d'un enfant de famille non éligible au programme). Les effets indirects du programme, autres que ceux qui affectent le revenu futur des bénéficiaires, ainsi que le taux d'activité sont ici ignorés. Le coût marginal (*i.e.*, pour un enfant supplémentaire) C des programmes préscolaires financés par le gouvernement est supposé être égal à :

$$C = C_0 + \varphi_h \Pr(D_i = h) + \varphi_c \Pr(D_i = c) - \tau p E(y_i) \quad (**)$$

Dans cette seconde expression, C_0 représente le coût fixe de gestion des programmes préscolaires, φ_h est le coût marginal d'une place supplémentaire dans le programme Head Start, φ_c est le coût marginal d'une place supplémentaire dans un autre programme préscolaire (*i.e.*, le programme « substitut »), $\Pr(D_i = h)$ est la probabilité pour un enfant i éligible de participer au programme *Head Start*, $\Pr(D_i = c)$ est la probabilité pour ce même enfant d'être scolarisé dans un autre programme préscolaire, et $\tau p E(y_i)$ est la recette fiscale provenant des revenus de cet enfant lorsqu'il sera salarié.

Supposons maintenant que la probabilité de recevoir une offre de place dans le programme Head Start augmente de δ points de pourcentage. Kline et Walters (2016) montrent que l'effet de cet accroissement sur la réussite scolaire mesurée par $E(y_i)$, noté $\partial E(y_i)/\partial \delta$, est égal au produit de la variation de la probabilité d'être incorporé dans le programme Head Start pour un enfant éligible mais préalablement non participant à ce programme (catégorie d'enfants qui était plus haut dénommée les « compliers »), notée $\partial \Pr(D_i = h)/\partial \delta$, par l'accroissement du score mesurant la réussite scolaire pour ce type d'enfant, notée $LATE_h$ ¹³². Mathématiquement, cela s'écrit :

$$\partial E(y_i)/\partial \delta = LATE_h \times \partial \Pr(D_i = h)/\partial \delta \quad (***)$$

Les deux termes de ce produit peuvent être observés ou calculés dans l'expérience HSIS qui a été présentée plus haut. En d'autres termes, si l'on souhaite calculer la variation de revenu due à un accroissement du nombre d'offres de places dans un programme, le protocole de l'expérimentation randomisée doit permettre d'observer ces deux composantes. Ce résultat dépend toutefois de deux hypothèses : 1) les offres de places dans le programme considéré sont distribuées aléatoirement, et 2) les utilités des familles éligibles ne dépendent pas de la variation δ de la probabilité de recevoir une offre de place dans le programme Head Start. Le bénéfice marginal d'un accroissement δ de cette probabilité est donc égal, en dérivant et en remplaçant (***) dans (*), à :

$$\partial B/\partial \delta = (1 - \tau)p \times LATE_h \times \partial \Pr(D_i = h)/\partial \delta$$

Le coût marginal correspondant est, en dérivant et en remplaçant (***) dans (**):

$$\partial C/\partial \delta = (\varphi_h - \varphi_c S_c - \tau p LATE_h) \times \partial \Pr(D_i = h)/\partial \delta$$

Dans cette expression, S_c représente la probabilité qu'un enfant participant précédemment à un programme scolaire « substitut » entre dans le programme Head Start. Le terme $\varphi_c S_c$ est donc le coût évité d'une participation au programme préscolaire substitut, cette somme se déduisant

¹³¹ Le paramètre p est le coefficient de proportionnalité introduit précédemment, à savoir celui qui permet de transformer le résultat scolaire en salaire de marché.

¹³² LATE est l'acronyme de « *local average treatment effect* », qui est la moyenne de l'effet dit « local » du programme (ou traitement) pour ceux des individus éligibles qui ne participaient pas préalablement au programme mais qui décident d'y participer lorsque l'offre leur en est faite.

du coût additionnel φ_h d'une place supplémentaire dans le programme Head Start. Le terme $\tau p \text{LATE}_h$ est la recette fiscale additionnelle associée à l'entrée dans le programme Head Start d'un enfant qui participait préalablement à un autre programme scolaire.

Le ratio des effets marginaux d'un accroissement marginal de la probabilité de recevoir une offre de place dans le programme Head Start est ce que Kline et Walters (2016) nomme la valeur marginale des fonds publics (notée MVPF pour « *marginal value of public funds* »). Ce ratio est ici égal à :

$$MVPF_{\delta} = \frac{\partial B / \partial \delta}{\partial C / \partial \delta} = \frac{(1 - \tau)p \times \text{LATE}_h}{(\varphi_h - \varphi_c S_c - \tau p \text{LATE}_h)}$$

Ce ratio représente la valeur d'un euro (ou d'un dollar) supplémentaire dépensé pour la mise en œuvre du programme, nette des recettes fiscales supplémentaires. C'est une métrique qui permet de comparer facilement l'efficacité de différents programmes concurrents. Son calcul ne nécessite pas que l'on spécifie l'origine des fonds finançant la mise en œuvre de chaque programme. Lorsque ce ratio est supérieur à 1, le programme est jugé comme efficace, même en présence de programmes substituts. L'évaluation n'est pas perturbée par la présence d'un programme substitut dès lors que le protocole de l'expérimentation permet d'identifier 1) la probabilité S_c qu'un enfant participant précédemment au programme scolaire substitut entre dans le programme Head Start ; et 2) l'accroissement, noté LATE_h , du score mesurant la réussite scolaire pour ce type d'enfant une fois qu'il a participé au programme Head Start.

Que retenir ?

Le traitement est la situation engendrée par la mise en application de l'investissement social. Le contrefactuel est la situation en l'absence de cet investissement.

L'évaluation de ses effets est rendue difficile car la situation contrefactuelle n'est pas observée pour les bénéficiaires, de même que le traitement ne l'est pas pour ceux qui n'en bénéficient pas.

L'expérimentation aléatoire contrôlée est considérée comme la procédure la plus à même de faire face à cette difficulté.

Elle peut elle-même être sujette à certains types de biais, au premier rang desquels figurent les biais d'adhésion (non compliance) et de substitution.

Le biais d'adhésion apparaît lorsque des individus éligibles, choisis aléatoirement pour participer au groupe cible, refusent d'y participer ou bien acceptent puis cessent de participer à l'expérience avant que celle-ci parvienne à son terme.

Le biais de substitution peut apparaître dès lors que certains membres du groupe de contrôle, n'ayant pas accès au programme durant le temps de l'expérience, cherchent à accéder à des substituts au programme avant la fin de l'expérience.

Ces deux biais peuvent toutefois être corrigés par l'introduction d'hypothèses peu contraignantes dans le cadre d'une expérimentation aléatoire contrôlée.

9. Recension et quantification des coûts de l'intervention

Les coûts liés à la mise en place de la politique alternative doivent être comparés à ceux de la situation contrefactuelle, qui prévaudrait en l'absence de mise en œuvre de cette politique. En règle générale, **il est préférable d'estimer le différentiel de coûts entre les deux options, mais une approche par le coût marginal peut se révéler supérieure dans certains cas, comme le montre l'exemple de la section précédente.**

Les coûts associés à la mise en œuvre d'une politique alternative sont parfois difficiles à évaluer. Il est toutefois important d'en faire au moins la liste et d'en proposer une approximation. La méthode dite « des ingrédients », développée par Levin et McEwan (2001)¹³³ et reprise par Dhaliwal *et al.* (2013),¹³⁴ est en ce domaine une référence utile (voir *infra*). Les chercheurs responsables de l'analyse devront faire le maximum, notamment en prenant contact avec les administrations et les agences compétentes, pour obtenir un chiffrage réaliste des coûts.

Malheureusement, les informations sur les coûts provenant de ces sources sont presque toujours incomplètes car les principes sur lesquels reposent les budgets n'ont pas été conçus pour donner des évaluations précises du coût économique. Les pratiques comptables conventionnelles, que ce soit dans le secteur des politiques sociales, éducatives, ou d'autres politiques publiques, ont été conçues à des fins autres que la prise en compte des coûts des interventions. Il est peu probable que les documents et rapports administratifs, ainsi que les sources budgétaires, rendent compte de toutes les composantes d'une intervention et de leurs coûts réels¹³⁵. En revanche, l'estimation des coûts à l'aide de la méthode des ingrédients s'appuie sur la connaissance qu'ont les évaluateurs des détails des interventions et elle utilise cette connaissance pour construire des estimations directes des coûts.

La méthode des ingrédients peut être décomposée en trois grandes étapes : 1) identifier et spécifier les « ingrédients » nécessaires à l'obtention des résultats de l'évaluation ; 2) déterminer leurs coûts, et 3) calculer les coûts totaux du programme et les coûts moyens par bénéficiaire. **Pour mener une analyse coûts-bénéfices ou coûts-efficacité, il est préférable d'intégrer la collecte de données sur les coûts en même temps que l'on évalue l'intervention.** La fiabilité des estimations peut être grandement améliorée lorsque les évaluations de coûts et d'efficacité sont menées simultanément.

La première étape d'une analyse des coûts consiste à identifier et à spécifier les ingrédients ou les ressources nécessaires à la mise en œuvre du programme évalué et à l'obtention du niveau d'efficacité attendu. **Ces ressources doivent être identifiées en détail, non seulement en termes quantitatifs, mais également qualitatifs.** Par exemple, les catégories de personnel doivent être précisées selon leurs qualifications, leurs fonctions et

¹³³ Levin H. et McEwan P. (2001), *Cost-Effectiveness Analysis*, Thousand Oaks, CA: Sage Publications Inc.

¹³⁴ Dhaliwal I., Duflo E., Glennerster R. et Tulloch C. (2013), "Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries: A General Framework with Applications for Education", dans *Education Policy in Developing Countries*, University of Chicago Press, p. 285-338.

¹³⁵ Cela plaide pour une systématisation des comptabilités analytiques dans les administrations porteuses d'investissement social, ou le perfectionnement de ces comptabilités lorsqu'elles existent.

leur implication en termes de temps. Un exercice similaire doit être effectué pour les installations, l'équipement et les autres intrants du programme, ainsi que pour les ressources nécessaires aux personnes bénéficiaires. Ces informations proviennent de trois sources : 1) de rapports descriptifs ; 2) d'observations provenant de l'intervention elle-même ; et 3) des entretiens avec le personnel gérant celle-ci. Les informations en provenance de ces différentes sources doivent être ensuite comparées afin de tester leur cohérence, les éventuelles divergences étant résolues par des investigations plus approfondies. Tous les ingrédients sont identifiés et spécifiés, quelle que soit la manière dont ils sont financés, c'est-à-dire qu'il est nécessaire d'identifier toutes les ressources qui contribuent au coût total, peu importe qui les a fournies ou acquises.

Une fois les ingrédients identifiés, l'étape suivante consiste à déterminer leurs coûts d'opportunité. Dans la mesure du possible, des prix de marché doivent être utilisés pour établir la valeur de chaque ingrédient. En ce qui concerne le personnel, les informations relatives aux quotités de temps et aux qualifications peuvent être associées à des prix de marché appropriés, tels que des salaires et des avantages monétaires. Les frais de personnel représentent généralement une grande partie des coûts des politiques d'investissement social. Une attention particulière doit donc leur être accordée.

Certains ingrédients peuvent être difficilement évalués à l'aide de prix de marché. C'est le cas par exemple d'un bâtiment appartenant à l'institution commanditaire et qui n'est pas utilisé dans la situation contrefactuelle ou des ressources en nature fournies par des bénévoles. Dans ces cas, on peut avoir recours à des prix fictifs. La valeur des ressources peut être en particulier estimée au moyen d'une procédure telle que l'évaluation contingente (voir, par exemple, Diamond et Hausman, 1994)¹³⁶, processus permettant de déterminer le prix qui serait offert pour acquérir ces ressources sur un marché concurrentiel. Les frais de personnel peuvent être obtenus à partir des lignes budgétaires correspondantes si elles existent, mais, dans la plupart des cas, il est conseillé d'utiliser comme référence les salaires moyens observés sur le marché du travail local ou national. Pour chaque catégorie d'ingrédients, il existe des méthodes pour déterminer les coûts afférents (voir Levin et McEwan, 2001, chapitre 3)¹³⁷. Dans le cas des établissements d'enseignement, les coûts annuels peuvent être calculés en identifiant les coûts de construction des nouvelles écoles et en amortissant ces coûts sur la durée de vie prévue de ces bâtiments.

Une fois les coûts des ingrédients estimés, la troisième étape consiste à agréger ces coûts pour déterminer le coût total du programme, annualisé ou évalué sur une période donnée à l'aide d'un facteur d'actualisation. Le coût moyen par participant et le coût marginal par participant doivent également être calculés. La mesure de coût la plus couramment utilisée est le coût moyen par participant, qui peut être combiné avec l'effet moyen par participant de façon à permettre une comparaison des rapports coûts-efficacité des différentes options possibles. Les ratios coûts-efficacité sont généralement calculés en divisant le coût par participant par l'efficacité moyenne observée. Les décideurs devraient privilégier les solutions les plus efficaces, *i. e.*, les moins coûteuses pour un niveau d'efficacité donné.

¹³⁶ Diamond P. et Hausman J. (1994), "Contingent Valuation: Is Some Number Better than No Number?", *Journal of Economic Perspectives*, Vol. 8(4), p. 45-64.

¹³⁷ Levin H. et McEwan P. (2001), *Cost-Effectiveness Analysis*. Thousand Oaks, CA: Sage Publications Inc.

Dans une dernière étape, il est possible d'évaluer la répartition des coûts entre les différents financeurs (par exemple, dans le cas d'une politique éducative, le ministère de l'Éducation, le Conseil régional ou départemental, la mairie, l'établissement scolaire, les parents d'élèves). Les contributions en nature, les subventions gouvernementales et les contributions philanthropiques peuvent être ici prises en compte. Ces ajustements permettent de calculer les coûts nets de l'intervention pour chaque type de financeur, public ou privé. Il est toutefois préférable de considérer les coûts de la manière la plus exhaustive possible, en y ajoutant le cas échéant des subventions reçues lors de la phase expérimentale ou pilote, si ces subventions peuvent être supprimées une fois le programme généralisé. Face aux incertitudes en la matière, l'évaluation gagnerait à comporter un jeu de chiffrages avec ou sans ces subventions.

Que retenir ?

Les coûts liés à la mise en place de la politique alternative doivent être comparés à ceux de la situation contrefactuelle, qui prévaudrait en l'absence de mise en œuvre de cette politique. Seuls les coûts différentiels sont à retenir.

Les coûts associés à la mise en œuvre d'une politique alternative sont parfois difficiles à évaluer. Il est toutefois important d'en faire au moins la liste et d'en proposer une approximation.

Lorsque les coûts ne sont pas directement observables, il est recommandé d'utiliser la méthode des ingrédients qui consiste en trois grandes étapes : 1) identifier et spécifier les ingrédients nécessaires à l'obtention des résultats de l'évaluation ; 2) déterminer leurs coûts, et 3) calculer les coûts totaux du programme et les coûts moyens par bénéficiaire.

Pour mener une analyse coûts-bénéfices ou coûts-efficacité, il est préférable d'intégrer la collecte de données sur les coûts en même temps que l'on évalue l'intervention.

Certains ingrédients peuvent être difficilement évalués à l'aide de prix de marché. Il est alors possible d'avoir recours à des évaluations contingentes.

Chapitre 4

Prolongements et spécificités du calcul socioéconomique appliqué à l'investissement social

Les chapitres précédents ont permis de définir les principales étapes et ingrédients du calcul socioéconomique appliqué aux investissements sociaux et de présenter des applications riches d'enseignements. Cependant, les enjeux de valorisation monétaires des effets, la prise en compte des effets redistributifs et d'équité et les effets d'équilibre général ont été peu évoqués jusqu'à présent. En réalité, il n'existe pas aujourd'hui de consensus qui permettrait de fournir des recommandations précises. Pour ces trois considérations, il existe une variété de méthodes et de pratiques plus ou moins adaptées aux situations. De plus, la mise en œuvre pratique d'évaluations d'investissements sociaux varie beaucoup suivant le type d'étude.

Les sections suivantes recensent et décrivent les méthodes visant à répondre aux enjeux de valorisation et d'équité d'une façon assez neutre, en faisant l'hypothèse qu'aucune de ces techniques n'est aujourd'hui véritablement meilleure que les autres.

1. Évaluations *ex-post* ou *ex-ante*, parcimonie ou complexité

L'évaluation socioéconomique d'un investissement social peut prendre des formes très différentes, selon qu'elle est réalisée en amont ou en aval de sa mise en œuvre. En outre, la prise en compte des effets plus ou moins directs des investissements sociaux se révèle généralement délicate, ces complexités conduisant finalement à une grande diversité d'analyses coûts-bénéfices. Dans l'annexe A, sont présentées en détail cinq évaluations qui reflètent des « cas types » d'analyses coûts-bénéfices. On les résume ici à grands traits dans l'optique de faire ressortir les principaux avantages et inconvénients qu'il y a à recourir à tel ou tel type de démarche d'évaluation.

1.1. L'évaluation *ex-post* parcimonieuse

Pour pouvoir mettre en œuvre l'évaluation *ex post* d'un investissement social, il importe de disposer au préalable d'évaluations d'impact permettant d'identifier des mécanismes causaux.

Pour évaluer l'effet d'une baisse de la taille des classes en fin de maternelle et à l'école primaire, Alan Krueger (1999) a utilisé les résultats de l'expérience randomisée STAR

conduite au Tennessee dans les années 1980. Il a ensuite proposé avec Whitmore¹³⁸ une évaluation coûts-bénéfices de cette politique.

Le calcul réalisé par Krueger et Whitmore (2001) est une application directe de la formule de la valeur actualisée nette (VAN) présentée au chapitre précédent. Ce calcul s'appuie sur une expérimentation aléatoire contrôlée qui fournit des estimations robustes des effets causaux et un suivi longitudinal long de cette intervention. Pour autant, cette approche reste parcimonieuse, car elle ne concerne que les effets sur les revenus et ne prend pas en compte l'ensemble des effets plus indirects qu'elle est susceptible d'avoir, tant du point de vue budgétaire (*via* les gains de recettes fiscales induits par les salaires plus élevés perçus par les bénéficiaires de la réduction de la taille des classes) que sociétal (en monétisant, par exemple, les effets de cette politique sur la qualité de la vie privée et familiale des bénéficiaires, leur santé, leur moindre propension à la délinquance, leur comportement citoyen, etc.). La valorisation monétaire des effets sur les revenus est fondée sur des études préalables (voir annexe) permettant de « convertir » des résultats scolaires en gains salariaux. Cette parcimonie délibérée tient au fait que les effets indirects ne sont pas directement observés et que leur prise en compte aurait dû reposer sur des hypothèses difficilement vérifiables.

La recherche de Krueger et Whitmore (2001) n'en est pas moins rigoureuse et instructive : elle permet en effet de tirer plusieurs enseignements méthodologiques pertinents :

- premièrement, elle est **adossée aux résultats d'une expérimentation randomisée qui produit des estimations robustes de l'effet moyen à court et moyen termes de la politique considérée** ;
- deuxièmement, elle **met en œuvre de façon particulièrement claire les principes du calcul de la variation de la valeur actualisée nette (VAN)** pour cette intervention, principes qui ont été exposés dans le chapitre précédent ;
- troisièmement, les **limites de l'exercice sont précisément énoncées, et chaque composante du calcul est soigneusement commentée**.

Les auteurs étant conscients des incertitudes pesant sur les valeurs possibles de ces composantes, ils **ne se bornent pas à un chiffrage unique** et considèrent plusieurs scénarios relativement plausibles.

Malgré toutes les précautions prises, visant à mettre en avant la part d'incertitude demeurant attachée à leur résultat, Krueger et Whitmore (2001) trouvent que, dans la plupart des scénarios envisagés, les bénéfices salariaux directs d'une réduction significative de la taille des classes sont plus élevés que les coûts de cette politique. De ce fait, les analyses coûts-bénéfices de Krueger et Whitmore (2001) sont souvent présentées comme « conservatrices » dans le sens où elles n'intègrent que les effets directs sur l'emploi. Cette restriction fournit une borne minimale au calcul coûts-bénéfices : dans la plupart des cas envisagés, réduire la taille des classes se trouve être un investissement socialement rentable.

¹³⁸ Krueger A. et Whitmore D. (2001), "The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR", *The Economic Journal*, vol. 111(468), p. 1-28.

Pour les évaluations *ex post* pour lesquelles il n'existe pas de données permettant de mesurer les effets induits potentiels, un modèle parcimonieux mais bien identifié est préférable à un modèle plus complet fondé sur des extrapolations. Même dans ce cas, des analyses de sensibilité sont indispensables pour asseoir la robustesse des résultats présentés.

1.2. Évaluations *ex ante* de politiques d'investissement social

Au-delà des intérêts inhérents aux études de type « *ex post* parcimonieux », il demeure souhaitable de produire un modèle complet visant à mesurer *ex ante* la valeur d'un investissement social. C'est ce que réalise le Washington State Institute for Public Policy (WSIPP). Il s'agit d'une organisation non gouvernementale et apolitique fondée en 1983 dans l'État de Washington aux États-Unis. Son but est de produire des recherches qui se veulent utiles et non partisans à destination des parlementaires et gouverneurs de l'État.

Le WSIPP réalise tout à la fois des analyses coûts-bénéfices des politiques actives sur le marché du travail, des politiques de lutte contre la délinquance, des politiques éducatives, de santé, d'accès à l'enseignement supérieur, etc., selon une méthodologie détaillée et identique pour toutes les politiques évaluées. Pour cela, il suppose que chacune de ces politiques doit être mise en place sur l'ensemble du territoire – c'est notamment à partir de cette hypothèse que le WSIPP évalue les coûts de mise en œuvre. Le WSIPP propose ainsi une méthodologie permettant de comparer les rendements estimés de différentes politiques.

Les évaluations ainsi produites sont exemplaires à bien des égards. De façon très simplifiée, le WSIPP procède en trois étapes :

- en premier lieu, il **recense et synthétise la littérature** mesurant les impacts de la politique examinée, puis calcule à partir de ce matériau les coefficients utilisés pour la méta-analyse ;
- les tailles d'effets précédemment calculées sont ensuite monétisées, projetées sur le cycle de vie puis comparées aux coûts par le biais du calcul de la valeur sociale nette actualisée et du ratio coûts-bénéfices ;
- le WSIPP estime enfin l'incertitude associée au calcul du rendement d'un investissement social à l'aide de simulations de Monte Carlo. L'objectif est ici d'estimer la probabilité que l'estimation de ce rendement soit positive.

L'usage de revues systématiques de littérature est une innovation majeure : au-delà du fait que des revues de ce type synthétisent l'ensemble de la connaissance sur un sujet à un moment donné, elles offrent bien d'autres avantages. En particulier, les coefficients des méta-analyses permettent de produire des distributions de valeurs quant aux effets de la politique considérée, donc un écart-type autour de la moyenne estimée de ces effets, qui peut ensuite être utilisé pour des calculs de sensibilité.

De plus, le WSIPP adopte une méthode permettant de « convertir » les effets sur différentes variables de résultats en une seule valeur monétisée. Pour ce faire, il construit une matrice de coefficients méta-analytiques reliant des résultats intermédiaires à d'autres variables qui ne sont généralement pas observées dans les enquêtes de suivi. Cette méthode lui permet

de réaliser une estimation des effets potentiels de long terme (non observés) à partir des effets observés sur un horizon plus court.

Il adopte également une méthode systématique pour éviter les doubles comptes (présentée en annexe). La valorisation monétaire des effets est fondée sur l'ensemble des évaluations économétriques fournissant des valeurs tutélaires pour les principaux effets considérés.

Les analyses de sensibilité sont ensuite produites à partir de simulations de Monte Carlo qui consistent à réestimer le ratio bénéfiques/coûts ou la valeur actualisée nette des centaines (voire des milliers) de fois, en faisant varier les paramètres du modèle. La méthode revient à tirer ces paramètres au hasard dans les distributions de probabilité des effets estimés. Une fois cet exercice terminé, le WSIPP calcule la moyenne des VAN simulées. Il en déduit une mesure d'incertitude qui est égale à la proportion de simulations pour lesquelles la VAN est positive.

Par la mise à disposition sur son site internet¹³⁹ de l'ensemble des revues de littérature et des applications produisant les estimations, le WSIPP facilite la compréhension et l'utilisation de ses travaux. En clair, si toutes les estimations qu'il produit peuvent être soumises à discussion, elles ont au moins le mérite de reposer sur un corpus d'hypothèses identifiables, éventuellement modifiables, et communes à chaque évaluation.

1.3. Scénario « What if » : estimer les effets sur la population

Les modèles d'évaluation *ex ante* du WSIPP sont d'une grande complexité mais ne produisent que des effets moyens ; ils ne tiennent pas compte des spécificités liées à la structure de la population, à son origine, etc. Ils ne rendent donc pas compte de l'hétérogénéité des effets dans la population. À l'opposé, le recours à la microsimulation peut permettre de rendre compte de **ce qui se passerait** (What if) dans une population ou un groupe donné si une ou plusieurs politiques étaient mises en œuvre, mais aussi de la façon dont leur rentabilité évoluerait au cours du temps.

En utilisant cette méthode, l'Américain Christopher Murray et l'Australien Alan Lopez ont réalisé un travail consistant à **évaluer le rapport coûts-efficacité de cent interventions préventives des maladies non transmissibles** et des facteurs de risque associés afin d'aider à déterminer celles qui sont les plus efficaces. Leur recherche a également produit l'analyse de cinquante traitements de diverses maladies non transmissibles et des moyens de lutte contre certaines maladies infectieuses. Le projet de recherche, qui a duré cinq ans, a donné lieu à la publication d'un rapport en 2010¹⁴⁰.

Dans cette étude, le cadre de l'analyse économique était limité aux dépenses de santé en Australie. La modélisation, qui est comparable à une microsimulation dynamique¹⁴¹, a été appliquée à la **population australienne de 2003 que les auteurs ont fait vieillir de quatre-vingts ans sans ajout de naissances**. Des analyses de sensibilité ont été

¹³⁹ <http://www.wsipp.wa.gov/>

¹⁴⁰ Vos T., Carter R., Mihalopoulos C., Veerman L., Magnus A. et Cobiac L. *et al.* (2010), *Assessing Cost-Effectiveness in Prevention*. Brisbane, Melbourne: University of Queensland and Deakin University, 101 pages.

¹⁴¹ Dans ce domaine, comme dans d'autres, on ne peut faire l'économie d'évaluations *ex ante* par le biais de simulations, et il est difficile d'éviter que celles-ci soient réalisées au niveau micro-économique.

réalisées : le principe était de faire varier les paramètres aléatoirement au sein de leurs marges d'incertitude pour vérifier la robustesse des résultats. Un taux d'actualisation de 3 % a été appliqué. Il s'agissait ici de comparer les pratiques préventives et de soins mises en place en 2003 ainsi que leurs coûts.

L'introduction dans le modèle des bénéfices des interventions a bien sûr comme limite la qualité des études réalisées et leur transférabilité à d'autres contextes. Ces facteurs ont été pris en compte en faisant appel à des groupes d'experts *ad hoc*, qui jugeaient, par exemple, que la qualité des données sur la prévention de l'obésité n'était pas satisfaisante, alors qu'elle l'était davantage dans le domaine de la prévention de la consommation de tabac et d'alcool.

Après ce travail analytique, les auteurs ont proposé à l'aide d'une modélisation appropriée la meilleure **combinaison de stratégies de prévention dans plusieurs domaines**, comme la prévention de la consommation d'alcool¹⁴², la prévention des risques solaires¹⁴³, la promotion de l'activité physique¹⁴⁴ ou la consommation de fruits et légumes¹⁴⁵.

1.4. Vers une évaluation *ex post* complète et complexe

Même dans une situation « idéale » où un investissement social est mis en place avec un protocole d'évaluation permettant de mesurer les effets de façon causale et capturant ces effets dans toutes les dimensions pertinentes, il reste des inconnues, des intervalles entre les dates d'observation qui font que, même dans ce cas favorable, on doit avoir recours à d'autres données ou à des hypothèses plus fortes.

C'est le cas par exemple du programme Perry Preschool, intervention phare qui a permis de proposer à des enfants issus de familles très défavorisées un accueil dans ce qui pourrait ressembler à des super-crèches en France, avec des professionnels très qualifiés, un projet pédagogique, etc. Dans le cadre de ce programme, chaque éducateur devait rendre visite à la famille de l'enfant une fois par semaine, environ pendant 90 minutes, afin d'évoquer avec les parents les progrès et les difficultés de l'enfant, la façon dont ils pouvaient l'accompagner dans le cadre du programme, mais aussi observer l'environnement familial de l'enfant. En résumé, ce programme pilote militait pour un investissement massif dans la petite enfance.

Il a été mis en place de façon expérimentale entre 1962 et 1967 dans la ville d'Ypsalanti (Michigan, USA) ; par tirage au sort, 58 enfants ont été aléatoirement affectés au groupe de traitement (ou groupe cible) et 65 enfants au groupe de contrôle.

Dès l'entrée à l'école élémentaire, les résultats scolaires des enfants du groupe cible étaient de loin supérieurs aux résultats des enfants du groupe de contrôle. Par la suite, les enfants du groupe cible ont été **plus longtemps scolarisés** (en moyenne pendant 11,9 années contre 11 années pour le groupe de contrôle), ils ont **passé moins de temps dans les**

¹⁴² Cobiac L., Vos T., Doran C. et Wallace A. (2009), "Cost-effectiveness of interventions to prevent alcohol-related disease and injury in Australia", *Addiction*, Vol. 104(10), p. 1646-55.

¹⁴³ Shih S., Carter R., Sinclair C., Mihalopoulos C. et Vos T. (2009), "Economic evaluation of skin cancer prevention in Australia", *Preventive Medicine*, Vol. 49(5), p. 449-53.

¹⁴⁴ Cobiac L., Vos T. et Barendregt J. J. (2009), "Cost-effectiveness of interventions to promote physical activity: a modelling study", *PLoS Medicine*, Vol. 6(7), e1000110.

¹⁴⁵ Cobiac L., Vos T. et Veerman J. L. (2010), "Cost-effectiveness of interventions to promote fruit and vegetable consumption", *PloS one*, Vol. 5(11), e14148.

établissements spécialisés, réservés aux enfants présentant des déficiences mentales ou émotionnelles, des difficultés d'élocution ou d'apprentissage (3,9 contre 5,2 années), ils ont **plus souvent obtenu un diplôme de l'enseignement supérieur** (66 contre 44 %). Dans le groupe cible, la proportion d'adolescentes enceintes était inférieure (0,6 contre 1,2 grossesse en moyenne). À 40 ans, les bénéficiaires du programme Perry Preschool ont été **moins souvent emprisonnés** (28 contre 52 %), ils percevaient un **revenu mensuel médian supérieur de 42 % à celui des membres du groupe de contrôle** (1 856 contre 1 308 dollars) et ils étaient moins nombreux à avoir bénéficié de l'aide sociale au cours des dix années précédentes (59 contre 80 %). **Toutefois, les critiques soulignent la petite taille de l'échantillon expérimental (123 enfants au total)**, l'absence d'effet substantiel à long terme sur le quotient intellectuel, et l'absence de significativité statistique de nombre de ses effets.

Pour évaluer le rendement du programme Perry Preschool, il faut pouvoir tenir compte de ce que : 1) le protocole de la randomisation a été compromis par le fait que certains enfants ont été réaffectés à l'un des deux groupes en dépit des résultats du tirage aléatoire préalablement réalisé ; 2) les individus ayant participé au programme ont été suivis jusqu'à l'âge de 40 ans, mais pas au-delà ; en conséquence, si l'on veut obtenir une estimation des bénéfices et des coûts du programme sur une période plus longue, par exemple jusqu'à 65 ans, il est nécessaire de recourir à des méthodes d'appariement avec d'autres sources statistiques ; 3) pour quelques individus n'ayant pas encore atteint 40 ans (*i. e.*, avant la fin de la période d'observation), certaines informations sont manquantes; et 4) il est difficile d'attribuer des valeurs monétaires fiables à certaines variables telles que les délits, dont la gravité peut être plus ou moins grande.

Dans un article publié en 2010,¹⁴⁶ Heckman et ses coauteurs ont proposé des solutions à chacun de ces problèmes et ont ensuite estimé le taux de rendement annuel moyen du programme Perry Preschool qui, selon eux, serait compris entre 7 % et 10 %. Par ailleurs, Heckman *et al.* (2010) ont calculé **les écarts-types de la plupart des effets estimés du programme**. Lorsque cela ne leur était pas possible, ils ont évalué la sensibilité de leurs résultats à certains paramètres qui ne peuvent pas être estimés, tels que le taux d'actualisation (qu'ils font varier de 0 % à 7 %), la « perte sèche » fiscale (« *deadweight loss* ») qui permet de financer le programme, ou encore la valeur monétaire d'une vie humaine (qui intervient dans le calcul du coût d'un meurtre). Heckman *et al.* (2010) ont utilisé plusieurs méthodes d'imputation pour compléter les informations manquantes (avant l'âge de 40 ans) ou celles qui ne peuvent être observées (entre 40 et 65 ans).

Cette évaluation semble être l'une des plus riches et des plus sophistiquées disponibles à ce jour. C'est un travail très important qui ne pourrait pas être facilement reproduit dans d'autres contextes. Elle doit être considérée comme ce qui peut se faire de mieux et servir de modèle pour les travaux futurs.

¹⁴⁶ Heckman J. J., Moon S. H., Pinto R., Savelyev P. A. et Yavitz A. (2010), "The rate of return to the High-Scope Perry Preschool Program", *Journal of Public Economics*, Vol. 94(1-2), p. 114-128.

2. Identifier la valeur sociale d'un résultat à l'aide d'un prix implicite

En principe, tous les effets de court et de plus long termes doivent être valorisés dans l'analyse et, autant que possible, distingués en fonction des types d'acteurs impliqués. Cependant, il existe une grande hétérogénéité de pratiques et de méthodes, à la fois dans la prise en compte réelle de ces effets et dans la façon de leur attribuer une valeur monétaire.

Les guides d'analyses coûts-bénéfices¹⁴⁷ proposent également d'autres méthodes, mais elles ne sont pas adaptées à l'évaluation des politiques d'investissement social. Par exemple l'utilisation de sous-produits pour estimer la valeur d'usage d'un bien public s'applique bien à la valeur d'une forêt – où l'on utiliserait le prix du bois –, mais plus difficilement à une politique sociale. Les méthodes difficiles à appliquer sont, selon nous, celles relatives à la mesure du coût d'évitement d'un effet donné ou celles concernant la limitation ou la restauration d'un résultat recherché.

Les sections suivantes présentent les méthodes généralement utilisées pour valoriser des coûts et bénéfices intangibles, c'est-à-dire des coûts non monétaires, qui ne sont généralement pas échangés sur un marché comme, par exemple, dans le cas des coûts de la délinquance, la valeur de la peur, de la douleur, de la souffrance, ou de la perte de qualité de vie.

2.1. *Shadow prices* ou prix implicites : identifier la valeur sociale d'un résultat

La notion de prix implicite est souvent utilisée par les économistes : elle décrit les situations dans lesquelles le prix de marché (s'il existe) ne reflète pas la véritable valeur du bien en raison des imperfections de marché, causées, par exemple, par la présence d'externalités, de biens publics, provoquant une distorsion de l'équilibre¹⁴⁸. Souvent même, il n'existe pas de marché du tout et la valeur sociale est inconnue. D'un point de vue conceptuel, le prix implicite d'un bien est défini comme le surplus social marginal, c'est-à-dire le gain ou la perte nette pour la société qui est associé au fait de bénéficier d'une unité en plus ou en moins de ce bien¹⁴⁹. Théoriquement, il s'agit d'une notion assez simple, mais dans la littérature empirique les prix implicites font surtout référence à des approximations de valeurs sociales obtenues à partir de préférences exprimées ou révélées.

La valeur économique associée à un bien, un service ou un résultat se décompose en une valeur liée à l'utilisation et une valeur associée à la non-utilisation de ce bien. Plus précisément, la valeur économique totale peut être divisée en trois composantes : sa valeur d'usage ; sa valeur d'option ; et sa valeur d'existence.

Un exemple permet de donner un sens plus précis à ces concepts : quelle valeur monétaire donner à la réduction du harcèlement scolaire au sein d'un établissement, école ou collège ?

La valeur d'usage est celle que l'on attribue au bénéfice retiré de l'utilisation du bien ou du service. Dans l'exemple ci-dessus, la valeur d'usage est celle qu'un parent attribue à

¹⁴⁷ Voir par exemple Romijn G. et Renes G. (2013), *General Guidance for Cost-Benefit Analysis*, CPB/PBL, The Hague; Treasury of New Zealand Government (2015), *Guide to social costs-benefits analysis*.

¹⁴⁸ Voir le chapitre "Shadow prices and market prices" dans l'ouvrage de Stiglitz J. (2000), *Economics of the public sector*, 3rd Edition, Columbia.

¹⁴⁹ Dreze J. et Stern N. (1990), "Policy reform, shadow prices, and market prices", *Journal of Public Economics*, Vol. 42(1), p. 1-45.

l'absence de harcèlement de son enfant. La valeur d'option est celle que l'on accorde à une utilisation potentielle ou future du bien. Dans notre exemple, il s'agit de la valeur qu'un parent attribue au fait que son enfant ne sera pas victime de harcèlement lorsqu'il sera scolarisé dans un établissement donné.

La valeur d'existence est celle que l'on attribue simplement au fait que le bien ou le service existe ou continue d'exister (ou l'inverse), indépendamment de son utilisation. Un parent peut ne retirer aucun bénéfice direct de l'absence de harcèlement scolaire dans l'école fréquentée par son enfant, mais il peut être farouchement opposé au harcèlement scolaire en général, auquel cas il valorisera la réduction du phénomène dans l'ensemble de la société.

D'autres notions de valeur sont parfois utilisées, comme celle de non-usage ou de préservation. Ces deux valeurs dont le nom est transparent renvoient à des concepts dont la définition et la classification divisent les économistes mais, même si leur définition fait débat, ceux-ci sont d'accord pour admettre que la seule méthode permettant de valoriser un effet non monétisé est celle de l'évaluation contingente¹⁵⁰.

2.2. Identifier le consentement à payer par le biais de l'évaluation contingente

La méthode de l'évaluation contingente utilise les préférences exprimées directement par les parties prenantes (usagers, citoyens, etc.). Elle se fonde sur les dispositions à payer et non pas sur les comportements de consommation réellement observés.

Collecter des données sur les préférences déclarées ou révélées consiste à demander à des personnes d'exprimer leur appréciation, sentiment, etc., à propos d'une action particulière et de ses effets potentiels. Les préférences déclarées sont une méthode d'évaluation contingente fondée sur des mesures du consentement à payer ou du consentement à recevoir afin de bénéficier ou non d'un bien ou d'un service, de se prémunir d'un risque, etc.

L'usage des valeurs obtenues par cette méthode peut varier suivant les situations. Elles peuvent être utilisées soit indépendamment d'une politique ou d'une mesure à évaluer, et viser une portée générale ; par exemple s'il s'agit de donner une valeur à une baisse de la délinquance ou du risque de cambriolage. Ou bien elles peuvent servir à estimer la valeur d'un investissement social dans son ensemble ; en identifiant le consentement à payer pour réduire son risque d'accident vasculaire lorsqu'on souffre d'hypertension.

Comme le rappellent Pearce *et al.*¹⁵¹, les notions de consentement à payer (CAP) et de consentement à recevoir (CAR) plongent leurs racines dans la théorie de l'économie du bien-être et correspondent aux concepts de variation compensatoire et de variation équivalente. En effet, dans un modèle classique où un consommateur maximise sous contrainte de revenu son utilité vis-à-vis d'un panier de biens, si les prix relatifs de ces biens sont modifiés, le montant de transferts de ressources qui permettrait de restaurer l'utilité initiale n'est pas le même suivant que l'on compense aux prix actuels (variation compensatoire) ou que l'on fournisse les niveaux de consommation correspondant aux prix initiaux. Le CAP et le CAR ne devraient pas, à en croire la théorie, être très différents l'un de l'autre. Il paraît pourtant exister entre eux un

¹⁵⁰ Kah E. (2003), « La méthode d'évaluation contingente appliquée aux déchets urbains », *L'Espace géographique*, Vol. 32(1), p. 47-59.

¹⁵¹ Pearce *et al.* (2006), *op. cit.*

écart parfois notable dans la pratique, le montant du CAR étant supérieur à celui du CAP. La décision de recourir au CAP ou au CAR n'est donc pas dénuée d'importance lors de la réalisation d'une analyse coûts-bénéfices¹⁵².

Un exemple hybride est fourni par Persson et Svensson (2013)¹⁵³. Ces deux économistes ont évalué le consentement à payer en vue de réduire les maltraitances et violences entre élèves au sein des établissements scolaires (*bullying*). Pour ce faire, ils ont mené une enquête auprès de la population d'une commune suédoise et ont évalué la disposition à payer de cette population à partir d'un scénario explicite. Celui-ci suppose que la municipalité prévoit de mettre en place un programme ayant pour but de réduire le nombre d'enfants victimes de violences à l'école, programme qui serait financé par une hausse des impôts locaux. Pour mesurer le consentement à payer des ménages, ils proposent cinq possibilités à chaque répondant : la réduction du risque et la variation du montant des impôts sont proposés au hasard et l'ordre des questions est également aléatoire. Les dispositions marginales à payer sont ensuite agrégées pour être comparées au coût de l'intervention. À partir des réponses obtenues, les auteurs identifient économétriquement le consentement à payer et les caractéristiques observables qui l'influencent. Le consentement marginal à payer pour qu'un élève ne soit pas harcelé à l'école est estimé en moyenne entre 66 cts et 85 cts d'euro. C'est-à-dire que tout calcul fait, chaque citoyen de la commune serait prêt à payer entre 66 et 95 cts de plus pour qu'un élève de moins soit préservé des violences. Cette donnée doit être ensuite agrégée pour être comparée au coût marginal de cet objectif. La population dans cette commune étant de 98 500 personnes, les auteurs estiment que la disposition globale à payer pour éviter une victime est comprise approximativement entre 65 000 et 93 000 euros.

Les auteurs calculent qu'un programme de ce type mis en place dans une école de 300 élèves coûte entre 11 000 et 46 000 euros par an. Les résultats de la littérature montrent que ce type de programme réduit de 20 % le nombre de victimes. Partant d'une situation où 30 élèves sont victimes de violences, 6 élèves environ seraient épargnés. Les bénéfices monétisés seraient alors compris entre 390 000 et 558 000 euros, soit un ratio coûts-bénéfices compris entre 8,4 et 50,7. Ce calcul pourrait toutefois être complété si l'on montrait qu'un programme de ce type a d'autres effets que celui de réduire le nombre d'élèves victimes de violences scolaires¹⁵⁴.

Ce type d'évaluations soulève deux types de problèmes conceptuellement importants. D'abord, parce qu'une personne valorise un bien, un service ou un résultat pour des raisons très différentes et qu'il n'est pas certain que ces raisons soient celles que cherchent à identifier ou mesurer le chercheur. L'évaluation contingente présuppose que les préférences énoncées par les personnes enquêtées sont bien celles que le chercheur souhaite connaître. Les personnes prennent parfois des décisions simplistes ou comprennent mal le protocole de l'expérience. En outre, dans l'exemple qui vient d'être présenté, il n'est pas certain que ce qui est ici mesuré reflète correctement la valeur du projet pour les citoyens – puisque c'est ainsi que la question est formulée – ou la valeur d'un élève qui, à la suite de la mesure, n'est plus harcelé.

¹⁵² Ce point a notamment été discuté par Nicolas Treich dans sa présentation au groupe de travail.

¹⁵³ Persson M. et Svensson M (2013), "The willingness to pay to reduce school bullying", *Economics of Education Review*, Vol. 35, p. 1-11.

¹⁵⁴ À condition toutefois de ne pas compter deux fois les effets. Si l'on considère par exemple une réduction de la délinquance à l'âge adulte, il faut se demander si ce résultat n'est pas la conséquence de la baisse de la violence en milieu scolaire, les harceleurs (ou victimes) pouvant devenir les délinquants de demain.

Par ailleurs, la valeur qu'une personne accorde à la réduction du harcèlement scolaire (d'une unité) peut être le résultat de normes sociales qui peuvent modifier le comportement des individus dans un sens qui n'est pas nécessairement celui attendu. Deux expériences souvent citées¹⁵⁵ montrent que la contractualisation d'un prix en vue d'inciter des individus à adopter un certain type de comportement peut produire l'effet inverse à celui recherché en modifiant le rapport de valeur et les motivations intrinsèques, ou bien celles liées à la réputation¹⁵⁶.

Le second problème est celui du **biais** des méthodes d'évaluation contingente. Celui-ci provient en premier lieu de la qualité de l'échantillon à partir duquel les paramètres sont estimés. Ces dernières peuvent être obtenues à l'aide d'un échantillon de trop petite taille ou d'un échantillon atypique, ou bien avec un taux de réponse faible¹⁵⁷. Ce peut être encore le cas lorsque les personnes interrogées sont particulièrement sensibles à la formulation ou à la nature des questions posées (des étudiants ou des parents seront probablement plus sensibles à la question du harcèlement scolaire que des adultes célibataires). En ce cas, les échantillons ne sont pas représentatifs. Il est par conséquent impératif d'avoir recours à un échantillon suffisamment grand et représentatif, mais aussi de s'assurer que le taux de réponse est satisfaisant. Mais ce problème n'est pas spécifique à l'évaluation contingente. En revanche, en demandant aux personnes de manière trop directe ce qu'est leur consentement à payer ou à accepter, on peut induire ces personnes à reporter une valeur inférieure ou supérieure à celle qu'elles estiment réellement. Cela peut être notamment le cas lorsque les situations proposées dans le questionnaire s'éloignent des situations auxquelles les répondants sont effectivement confrontés ; bien souvent, les questions posées décrivent un ensemble incomplet de possibles qui contraint les répondants à se positionner face à des contingences peu probables. Les évaluations contingentes peuvent également souffrir d'un biais stratégique lorsque les personnes interrogées manipulent les réponses afin d'orienter les résultats de l'enquête dans le sens de leurs propres intérêts, ou au contraire lorsqu'elles se comportent en *passagers clandestins*¹⁵⁸, laissant aux autres la responsabilité des réponses.

Il existe néanmoins des mécanismes d'enchères qui incitent les répondants à révéler leur véritable disposition à payer et à limiter fortement les comportements stratégiques. Bien que la loterie soit une procédure peu pertinente dans le cas de la valorisation des effets des politiques sociales, celle proposée par Becker *et al.*¹⁵⁹ (appelée par la suite mécanisme BDM) peut être utilisée pour révéler le consentement à payer pour un bien ou un service public. Le mécanisme BDM est simple : les participants sont invités à proposer une valeur puis un prix est tiré au hasard. S'il est inférieur à cette valeur, les participants payent le prix aléatoirement offert et

¹⁵⁵ U. Gneezy et A. Rustichini ont montré que l'introduction d'une amende donnée aux parents qui arrivent en retard pour récupérer leurs enfants en crèche a eu pour effet d'augmenter le nombre de retards : Gneezy U. et Rustichini A. (2000), "A fine is a price", *Journal of Legal Studies*, Vol. 29(1), p. 1-18. Dans le même ordre d'idées, N. Lacetera et M. Macis ont montré que la mise en place d'une compensation monétaire offerte aux donneurs de sang a réduit le nombre de dons : Lacetera N. et Macis M. (2013), "Time for blood: The effect of paid leave legislation on altruistic behavior", *Journal of Law, Economics, and Organization*, Vol. 29(6), p. 1384-1420.

¹⁵⁶ Tirole J. (2009), « Motivation intrinsèque, incitations et normes sociales », *Revue économique*, Vol. 60(3), p. 577-589.

¹⁵⁷ Dans l'expérience précédemment décrite, le taux de réponse était inférieur à 40 %.

¹⁵⁸ Le Gall-Ely M. (2009), « Définition, mesure et déterminants du consentement à payer du consommateur : synthèse critique et voies de recherche », *Recherche et Applications en Marketing* (French Edition), SAGE Publications, Vol. 24(2), p.91-113.

¹⁵⁹ Becker G., DeGroot M. et Marschak J. (1964), "Measuring utility by a single-response sequential method", *Behavioral Science*, Vol. 9(3), p. 226-232.

bénéficient alors de la mesure de politique publique, sinon l'offre est rejetée. La valeur initiale apparaît ainsi comme la somme limite que les participants sont prêts à payer, soit le consentement à payer. Ce mécanisme a été mis en œuvre au Ghana afin de mesurer le consentement à payer pour de l'eau saine dans le cadre d'un programme proposant des filtres à eau¹⁶⁰. Dans le cas où l'investissement social considéré n'est pas complètement gratuit, ce type de mécanisme peut permettre à la fois de mesurer sa valeur sociale et de fournir une source d'identification de son effet sur les bénéficiaires en raison du tirage aléatoire du prix¹⁶¹. Cependant, bien que ce mécanisme fournisse des incitations permettant de révéler le véritable consentement à payer, il n'élimine pas entièrement les comportements stratégiques, car il est possible que les agents sous-estiment la valeur du bien s'ils sont très adverses au risque, ou inversement la surestiment¹⁶².

2.3. Identifier la valeur sociale par les préférences révélées

Alors que les évaluations contingentes cherchent à expliciter les préférences individuelles, les méthodes dites des préférences révélées reposent sur l'observation des comportements face à des situations complémentaires ou substituables afin d'en déduire la valeur sociale d'un bien, d'un service ou d'un résultat.

Une première méthode est celle dite du *marché analogue* : pour mesurer la valeur accordée à un service public, il est possible d'observer le prix du même service (ou tout au moins comparable) sur un marché (privé), à condition que celui-ci soit concurrentiel. Par exemple, si l'on veut caractériser la valeur sociale associée à la réussite éducative, ou inversement celle associée à l'évitement d'un échec scolaire, on peut utiliser les prix pratiqués sur le marché des cours de soutien privé et les dépenses auxquelles les familles consentent en vue de la réussite scolaire de leur enfant. Cependant, la demande sur ce marché peut être jugée très atypique et ne pas refléter les préférences de toute la société¹⁶³. Une autre méthode est celle dite des *choix alternatifs* (ou *trade-off*) qui consiste à évaluer la valeur d'un résultat par son coût d'opportunité. L'exemple le plus souvent pris est celui du temps consacré à étudier, ou la valeur du temps de loisir, qui peut être évalué par le montant du salaire qui aurait été perçu si la personne avait consacré ce temps d'étude ou de loisir au travail rémunéré. À l'évidence, cette méthode est potentiellement biaisée. Dans le cas de ce dernier exemple, le temps de travail est supposé être flexible, alors qu'en réalité ce n'est généralement pas le cas. En outre, cette approche ne tient pas compte des avantages en nature liés à l'emploi, des impôts et autres charges payées par les salariés, ou bien encore de la possibilité d'occuper plusieurs emplois différents simultanément.

La méthode des *prix hédoniques* peut être également utilisée pour valoriser certains types d'effets. Le principe de cette méthode est d'utiliser la valeur d'un bien ou d'un service de

¹⁶⁰ Berry J., Fischer G. et Guiteras R. (2012), Eliciting and utilizing willingness to pay: Evidence from water filters in Ghana, International Growth Center.

¹⁶¹ Chassang S., Padro G., Miquel I et Snowberg E. (2012), "Selective trials: A principal-agent approach to randomized controlled experiments", *American Economic Review*, vol. 102(4), p. 1279-1309.

¹⁶² Kaas K. P. et Ruprecht H. (2006), "Are Vickrey auction and the BDM mechanism really incentive compatible? Empirical results and optimal bidding strategies in cases of uncertain willingness to pay", *Smalenbach Business Review*, Vol. 58(1), p. 37-55.

¹⁶³ Voir Galinié A. et Heim A. (2016), « Inégalités scolaires : quels rôles jouent les cours privés ? », Contribution au rapport du Cnesco *Les inégalités scolaires d'origines sociales et ethnoculturelle*, Paris, Cnesco.

substitution pour mesurer la valeur implicite d'un bien ou service pour lequel n'existe pas de marché. Par exemple, le prix de l'immobilier peut être utilisé pour attribuer une valeur à un bien public. La méthode des prix hédoniques consiste à régresser le prix d'un bien sur l'ensemble des caractéristiques observables susceptibles de l'affecter (dans le cas d'un bien immobilier, cet ensemble regrouperait la superficie, le nombre de pièces, l'ancienneté de la construction, etc.) et d'estimer par ce moyen la valeur du coefficient associé à la caractéristique considérée. Cette technique permet par exemple d'évaluer le prix d'une crèche publique. Sa mise en œuvre suppose toutefois que les individus aient au préalable une bonne connaissance des externalités possibles, disposent d'informations précises sur les caractéristiques des logements dans le quartier, etc. Par ailleurs, plusieurs des variables explicatives (*i.e.*, les caractéristiques du logement) utilisées dans ces modèles hédoniques sont souvent colinéaires, si bien qu'une fois intégrées dans le modèle explicatif, elles diminuent significativement la précision des coefficients estimés.

Les données recueillies par le système judiciaire et les statistiques publiées par les compagnies d'assurance pourraient également aider à construire certaines valeurs tutélaires. Les juges sont amenés à définir les dommages et intérêts pour des préjudices de natures variables, mais à notre connaissance ces données n'ont jamais été exploitées dans ce sens en France. Une telle exploitation pourrait au demeurant être entravée par un manque de cohérence et/ou de stabilité de la jurisprudence. De façon similaire, les compagnies d'assurance définissent des montants d'indemnisation dans le cas de nombreux types de dommages, et ces montants pourraient être exploités pour établir des valeurs tutélaires.

En l'absence d'autres moyens indirects pour estimer la valeur d'un effet induit, il est possible d'avoir recours à des *valeurs tutélaires fondées sur des dires d'experts*¹⁶⁴. Par exemple, la *méthode Delphes (Delphi method)*, développée par Dalkey et Helmer (1963)¹⁶⁵, permet d'obtenir le plus grand consensus possible à partir des évaluations d'un panel d'experts interrogés sur des sujets complexes et incertains. Il s'agit d'un processus structuré et interactif qui facilite la communication et la prise de position au sein d'un groupe en systématisant les échanges d'avis éclairés¹⁶⁶. La méthode s'est montrée particulièrement efficace pour traiter des situations dans lesquelles les données sont rares ou peu fiables, ou dans lesquelles l'incertitude est grande. Les experts interrogés doivent commencer par fournir des premières valeurs pour les effets considérés avant de les confronter et de les discuter systématiquement au sein du groupe ; chacun peut ensuite les modifier ou non à la suite de cette discussion collective. Le processus est itéré jusqu'à ce qu'un large consensus émerge. Ramos *et al.* (2016)¹⁶⁷ ont adopté la méthode Delphes afin d'intégrer les externalités dans les analyses relatives à la prévention des risques professionnels. La mise en place de ces politiques ou

¹⁶⁴ Un exemple particulièrement utile dans le cas français est [le tableau de valeurs tutélaires](#) fourni par le groupe d'experts qui a rédigé, sous la direction de Roger Guesnerie, le *Guide de l'évaluation socioéconomique des investissements publics* (2017). Ce tableau de valeurs tutélaires couvre les champs de la santé, du climat, de l'environnement, du temps et du bruit. Pour évaluer des politiques d'investissement social, il est fortement recommandé d'utiliser certaines des valeurs fournies dans ces tableaux, notamment celles relatives à la santé (valeur statistique de la vie humaine, valeur de l'année de vie, valeur du blessé léger, valeur du blessé grave) et au temps (déplacement professionnel, déplacement domicile-travail/études/garde, etc.).

¹⁶⁵ Dalkey N. C. et Helmer O. (1963), "An experimental application of the Delphi Method to the use of experts", *Management Science*, Vol. 9(3), p. 458-468.

¹⁶⁶ Linstone H. et Turrof M. (1975), *The Delphi method, techniques and applications*, Addison Wesley Publishing.

¹⁶⁷ Ramos D., Arezes P. et Afonso P. (2016), "Application of the Delphi Method for the inclusion of externalities in occupational safety and health analysis", *DYNA*, Vol. 83(196), p. 14-20.

dispositifs peut en effet avoir de nombreux autres effets que ceux observés au sein des entreprises ; il est alors nécessaire de prendre en compte la valorisation de ces effets externes dans le cadre d'une analyse coûts-bénéfices complète.

2.4. Les coûts des délits aux États-Unis : illustration d'un problème complexe.

Le coût d'un délit est un exemple particulièrement pertinent permettant d'illustrer la prise en compte de coûts tangibles et intangibles, mais aussi parce qu'il illustre certains des problèmes conceptuels inhérents à la monétisation de valeurs sociales.

Cohen¹⁶⁸ (2000) pose la question de savoir si l'on doit considérer la valeur d'un délit sous l'angle de l'externalité qu'il représente ou du coût qu'il fait peser sur la société. D'après lui, le coût externe est le coût imposé à une personne par une autre dans le cas où la première n'a pu volontairement accepter les conséquences. Par exemple, dans le cas d'un vol, les coûts externes concernent les biens volés, les coûts des soins, la perte de salaire, la souffrance et la douleur de la victime. Par ailleurs, la société estime que faire supporter de tels coûts à une victime est moralement répréhensible et contre la loi, cela justifiant la valorisation. Le concept de coût social est lié mais pas identique. En effet, le coût social correspond aux coûts qui affectent le bien-être agrégé de la société. Dans cet exemple, le coût des soins, la perte de salaire sont clairement une perte sèche pour la société et, même s'il n'y a pas de marché pour cela, on peut s'attendre à ce que la société consente à payer pour éviter à ses membres la douleur et la souffrance associées au vol. Ainsi, à partir du moment où la société considère le bien-être des victimes de délits, ces coûts font partie des coûts sociaux. En revanche, la valeur des biens volés pose un problème. En effet, les biens volés n'entrent techniquement pas dans les coûts sociaux puisque le voleur tire une utilité de leur utilisation ou de leur revente.

Quel coût considérer alors ? Implicitement, intégrer la valeur des biens volés revient à exclure du bien-être collectif les personnes ayant commis des infractions ; leur bien-être doit être ignoré de la fonction de bien-être social, ce que propose explicitement Trumbull¹⁶⁹ (1990). McChesney¹⁷⁰ (1993) considère pour sa part que les activités criminelles sont assimilables à une recherche de rente et n'ont pas de valeur sociale. Ce débat académique illustre combien la question des coûts des délits est une question normative et qu'il n'y a pas aujourd'hui de consensus sur la façon de les estimer.

Indépendamment de la question du coût social ou externe, il est nécessaire de recouvrer dans tous les cas les coûts tangibles d'un côté, et les coûts intangibles de l'autre.

Si, à première vue, l'évaluation des coûts tangibles peut paraître simple, celle-ci est en réalité un exercice compliqué, à l'exception des données sur les dépenses publiques supportées par le système judiciaire qu'il est plus facile de collecter. Par exemple, Cohen rappelle qu'il n'y a aucune comptabilité de la valeur des biens perdus par la victime d'un délit. Par ailleurs, s'il est généralement possible d'obtenir des données agrégées sur les dépenses et les coûts relatifs

¹⁶⁸ Cohen M. A. (2000), Measuring the Costs and Benefits of Crime and Justice, *Criminal Justice*, janvier.

¹⁶⁹ Trumbull et William N. (1990), "Who has standing in cost-benefit analysis?", *Journal of Policy Analysis and Management*, 9:201-218, cité dans Cohen (2000).

¹⁷⁰ McChesney et Fred S. (1993), "Boxed in: Economists and the benefits of crime", *International Review of Law and Economics* 13 (2): 225-231.

aux procédures judiciaires, il n'est généralement pas possible d'estimer le coût d'un délit de type donné.

Il est donc nécessaire de s'engager dans un exercice de comptabilité analytique, de répartir les coûts d'utilisation du système judiciaire au prorata des interventions, etc.

Par la suite, cette valorisation est susceptible d'évoluer dans le temps. Une façon de calculer le coût tangible moyen pour chaque délit est d'utiliser les techniques d'analyse de séries temporelles et d'identifier soit des relations de cointégration, soit de façon beaucoup plus fruste de décomposer l'évolution des coûts en fonction de l'évolution des délits à l'aide de modèles linéaires. C'est ce que fait le WSIPP, en utilisant également les travaux de Cohen pour évaluer les valeurs intangibles.

Cost category	Party who directly bears cost*	Cost category	Party who directly bears cost*
Direct property losses		Precautionary expenditures/effort	Potential victim
Losses not reimbursed by insurance	Victim	Fear of crime	Potential victim
Losses reimbursed by insurance	Society	Criminal justice system	
Administrative cost: Insurance reimbursement	Society	Police and investigative costs	Society
Medical and mental health care		Prosecutors	Society
Charges not reimbursed by insurance	Victim	Courts	Society
Charges reimbursed by insurance	Society	Legal fees	
Administrative overhead of insurance	Society	public defenders	Society
Victim services		private	Offenders
Expenses charged to victim	Victim	Incarceration	Society
Expenses paid by agency	Society	Nonincarcerative sanctions	Society
Temporary labor and training of replacements	Society	Victim's time	Victim
Lost workdays		Jury's and witness' time	Jury/witness
Lost wages for unpaid workdays lost	Victim	Victim services	
Lost productivity for paid workdays	Society	Victim service organizations	Society/volunteers
Lost schooldays		Victim compensation programs	Society
Foregone wages due to lack of education	Victim	Victim's time	Victim
Foregone nonpecuniary benefits of education	Victim	Other noncriminal programs	
Foregone social benefits due to lack of education	Society	Hotlines and public service announcements	Society/volunteers
Lost housework	Victim	Community treatment programs	Society
Pain and suffering/quality of life	Victim	Private therapy/counseling	Society/offender
Loss of affection/enjoyment	Victim's family	Neighborhood watch and community prevention programs	Volunteers
Death		Incarcerated offender	
Value of life	Victim	Lost wages	Offender/family
Funeral and burial expenses	Victim's family	Lost tax revenue and productivity	Society
Loss of affection/enjoyment	Victim's family	Value of lost freedom	Offender
Psychological injury/treatment	Victim's family	Psychological cost to family	Family of offender
Legal costs associated with tort claims	Victim or victim's family	"Overdeterrence"	
Long-term consequences of victimization	Future victims and society	Innocent individuals accused of offense	Innocent individuals
		Restriction of legitimate activity	Innocent individuals/society
		Cost of detection avoidance by offenders	Offender/society/victim
		"Justice"	
		Constitutional protections to avoid false accusations	Society
		Increasing detection to avoid differential punishment	Society

* Ignores any recovery from offenders through legal action.

Source: Cohen, Miller, and Rossman 1994.

* Ignores any recovery from offenders through legal action.

Source: Adapted from Cohen, Miller, and Rossman 1994.

Pour ce qui est des coûts intangibles, les méthodes décrites dans ce chapitre ont toutes été utilisées par différents auteurs : évaluations contingentes, préférences révélées (via les prix de l'immobilier dans les quartiers plus sécurisés) et dédommagements décidés par les jurys d'assises. C'est généralement cette partie des coûts qui est la plus élevée, la plus incertaine et la plus controversée. Wickramasekera *et al.*¹⁷¹ (2015) ont réalisé une revue de littérature systématique estimant les coûts des délits.

Sur les 21 études qui satisfont leurs critères d'inclusion, les estimations varient considérablement d'un pays à l'autre, mais aussi pour le même pays. Ainsi aux USA, où les coûts sont très largement supérieurs à ce qu'ils sont dans les autres pays, la valeur sociale d'un homicide va de 6,9 à 17 millions de dollars 2008 dans deux études publiées la même année.

Dans les calculs socioéconomiques d'investissements sociaux, la façon de monétiser les coûts des crimes et délits évités dépend par ailleurs beaucoup de la qualité des informations contenues dans les ensembles de données individuelles permettant de suivre les bénéficiaires des politiques considérées. Plus ces informations sont précises, par exemple parce qu'elles détaillent les types de délits commis ou les durées des périodes durant lesquelles les personnes ont perçu des minima sociaux, plus facile et précise est leur monétisation.

Prenons l'exemple du cas des délits commis par les personnes ayant participé au programme Perry Preschool, tel qu'il est évalué par Heckman *et al.* (2010). Les données permettent de connaître avec précision le nombre d'arrestations, d'accusations, de condamnations et d'incarcérations de chaque participant au programme (jusqu'à l'âge de 40 ans). Pour estimer les coûts correspondants, Heckman *et al.* (2010) utilisent différentes sources administratives : le *Uniform Crime Report* (UCR) publié par le Federal Bureau of Investigation, le National Crime Victimization Survey (NCVS) et les données individuelles de l'Expenditure and Employment Data for the Criminal Justice System (CJEE). Les coûts supportés par les victimes sont décrits par Cohen (2005)¹⁷² dans son ouvrage sur le coût du crime et de la justice. Les coûts d'incarcération sont calculés à partir d'informations fournies par les établissements pénitentiaires du Michigan.

Face à ces résultats, il apparaît donc nécessaire de mener un vaste projet d'évaluation des coûts de la criminalité pour la société, ce programme de travail devant intégrer à la fois une revue systématique des travaux passés, la mise en place d'un panel d'experts visant à déterminer les coûts à prendre en compte, et la mise en place d'évaluations contingentes afin d'estimer les coûts intangibles.

¹⁷¹ Wickramasekera N., Wright J., Elsey H., Murray J. et Tubeuf S. (2015), "Cost of Crime: A systematic review", *Journal of Criminal Justice*, 43 (3). 218 - 228. ISSN 0047-2352

¹⁷² M. A. Cohen (2005), *The Costs of Crime and Justice*, Routledge, New York.

Que retenir ?

La valeur sociale d'un effet est associée à la notion de prix implicite défini comme le surplus social marginal, c'est-à-dire le gain ou la perte nette pour la société, associé au fait de bénéficier d'une unité en plus ou en moins de ce bien. Dans la littérature empirique, les prix implicites font surtout référence à des approximations de valeurs sociales obtenues à partir de préférences exprimées ou révélées.

La valeur économique associée à un bien, un service ou un résultat se décompose en une valeur liée à l'utilisation et à la non-utilisation de ce bien.

Une méthode pour mesurer la valeur sociale consiste à utiliser l'évaluation contingente, c'est-à-dire des enquêtes permettant d'obtenir les préférences exprimées directement par les parties prenantes, en particulier le « consentement à payer » et le « consentement à recevoir ».

Les évaluations contingentes souffrent de plusieurs biais qui rendent leur utilisation délicate, même si en théorie ce sont précisément ces quantités que l'on cherche à estimer.

Une autre façon d'estimer les valeurs sociales consiste à utiliser les préférences révélées, c'est-à-dire des valeurs ou prix qui reflètent indirectement ces préférences. Ces méthodes indirectes sont par définition des approximations et leur validité est souvent remise en cause.

Face à ce constat, il apparaît nécessaire d'entreprendre un vaste projet d'évaluation des coûts sociaux, dont le périmètre inclurait tout à la fois une revue systématique des travaux disponibles, la mise en place d'un panel d'experts visant à déterminer les coûts à prendre en compte, et la mise en place d'évaluations contingentes afin d'estimer les coûts intangibles.

3. Équité et redistribution dans les analyses coûts-bénéfices

La plupart des analyses coûts-bénéfices produites à l'échelle internationale somment les bénéfiques et les coûts sans nécessairement se préoccuper de savoir s'il existe des groupes perdants en raison de la mise en œuvre du projet. Traditionnellement, la justification de cette absence de prise en compte revient à la possibilité de compenser les perdants selon le principe de Hicks et Kaldor. En théorie, les bénéfices nets d'une politique d'investissement social efficace seraient suffisants pour compenser potentiellement les perdants de cette mesure sans que personne d'autre ne soit lésé [car un investissement social « rentable » correspond à une amélioration au sens de l'équilibre de Pareto¹⁷³]. S'il est possible de compenser les perdants sans perte sèche pour la société – le coût de la compensation restant inférieur ou égal à la valeur sociale nette actualisée –, alors la société dans son ensemble est gagnante, ce qui est considéré comme suffisant.

¹⁷³ Pol T. V., Bos F. et Romijn G. (2016), "Distributionally weighted cost-benefit analysis: From theory to practice", *CBP discussion paper 364*, Den Haag.

Cependant, lorsque l'on estime le rendement d'un investissement social tel qu'une politique en faveur de la petite enfance, une politique active sur le marché du travail ou un dispositif de prévention de la délinquance juvénile, la redistribution et la justice sociale sont souvent des objectifs en soi, présentés comme cibles explicites de la politique¹⁷⁴.

Les enjeux redistributifs sont relatifs à l'identification des gagnants et des perdants. La question de la redistribution concerne la répartition des coûts et des gains entre des groupes spécifiques (bénéficiaires de la politique, non bénéficiaires, contribuables, etc.) ou des groupes qui présentent un intérêt particulier pour le décideur (par exemple, des groupes de personnes – ou de ménages – définies par leur niveau de revenu, leur origine sociale ou géographique, leur âge, etc.). Les enjeux d'équité ont trait à des questions plus normatives et interrogent le caractère juste et équitable des retombées ou des coûts d'une politique publique. La prise en compte de ces considérations dans l'analyse elle-même peut modifier la valeur sociale nette actualisée qui est habituellement calculée.

La littérature sur les méthodes de prise en compte de l'aversion aux inégalités dans les analyses coûts-bénéfices est croissante¹⁷⁵. **La plupart de ces méthodes impliquent de pondérer les gains et les pertes de certains groupes au moyen d'une fonction de bien-être sociale incluant un objectif d'égalisation de la répartition de certains *outcomes* dans la population.** Toutefois ces travaux restent encore théoriques et les méthodes proposées ont rarement fait l'objet d'applications dans des contextes informationnels proches de celui de la décision publique. De plus, l'utilisation de telles pondérations dans le cadre des analyses coûts-bénéfices devrait aller de pair avec l'organisation de discussions explicites sur les principes redistributifs que la collectivité souhaiterait concrétiser.

À côté de cette littérature théorique, on constate que les pratiques actuelles concernant l'intégration des enjeux d'équité dans l'analyse coûts-bénéfices relèvent le plus souvent d'une démarche *ad hoc*, subjective et, en tant que telle, non consensuelle. Les guides les plus complets recommandent généralement de s'en tenir à une répartition des effets¹⁷⁶. Le problème provient tout à la fois de la définition des pondérations et de la façon de les intégrer dans le calcul. Certaines pondérations explicites sont plus controversées que d'autres. Par exemple, en 2003, l'agence de protection de l'environnement des États-Unis (*U.S. Environmental Protection Agency*) a évalué un projet en accordant une valeur statistique à la vie humaine dépendant de l'âge : la valeur d'une vie d'une personne de plus de 65 ans était estimée inférieure de 37 % à celle d'une personne âgée de 16 à 64 ans. Cette pondération a suscité une telle controverse¹⁷⁷ que l'agence a dû opter pour une valeur tutélaire unique pour la vie humaine.

¹⁷⁴ Weimer D. et Vining A. (2010), *Policy Analysis: Concepts and Practice*. Routledge, Oxford.

¹⁷⁵ Voir Adler M.D (2011), *Well-Being and Fair Distribution: Beyond Cost-Benefit Analysis*, Oxford University Press ; Fleurbaey M. et Adler M.D. (2016), *The Oxford Handbook of Well-Being and Public Policy*, Oxford University Press.

¹⁷⁶ C'est par exemple le cas du guide du trésor de Nouvelle-Zélande (2015, *op. cit.*), de l'Australian Department of Finance Handbook of Cost Benefit Analysis (2006) et de celui du CPB (2016) hollandais. *A contrario*, le *UK Treasury Green Book* recommande d'utiliser des pondérations redistributives sans pour autant donner d'indications techniques ou pratiques sur la façon de le faire.

¹⁷⁷ Voir notamment Aldy J. E. et Viscusi W. K. (2015), "Labor market estimates of the senior discount for the value of statistical life", *Journal of Environmental Economics and Management*, 53 (2007) 377-392.

Kriström¹⁷⁸ (2005) indique trois façons de tenir compte des enjeux d'équité dans une analyse coûts-bénéfices. La première consiste à identifier et à présenter la manière dont se répartissent les coûts et les bénéfices liés aux projets ; la seconde, plus élaborée, consiste à calculer des coefficients implicites de pondération (par exemple, si un projet génère des pertes nettes globales, mais qu'un groupe de personnes dont la société ou l'autorité politique se soucie particulièrement en retire des gains nets, quel devrait être le coefficient attaché à ce groupe pour que la valeur pondérée du projet soit égale à 0 ?) ; la troisième consiste explicitement à recalculer les bénéfices nets du projet, sur la base de coefficients explicites de pondération, en fonction de la répartition des bénéfices reçus et des coûts supportés par les différents groupes sociaux.

Plusieurs économistes s'opposent à tout usage des évaluations coûts-bénéfices qui ne se baseraient que sur le critère d'optimalité de Hicks et Kaldor. Ainsi, Manski¹⁷⁹ (2015) plaide en faveur de l'utilisation de pondérations systématiques pour chaque membre de la société au motif que l'amélioration de la situation de chacun doit être l'objectif de toute politique et que les mécanismes de redistribution doivent être pensés en amont et ne pas être une simple possibilité.

Enfin la prise en compte des enjeux d'équité dans l'analyse coûts-bénéfices interroge également le choix de la perspective de l'évaluation *ex ante* ou *ex post*. D'une part, il apparaît que l'intégration de différents degrés d'aversion aux inégalités a plus d'impact sur le résultat des évaluations lorsqu'une perspective *ex post* est retenue¹⁸⁰, comme cela est discuté dans la section **Erreur ! Source du renvoi introuvable.** D'autre part, le choix de la perspective peut renvoyer à des positions normatives, par exemple sur la question de la responsabilité¹⁸¹.

La valorisation monétaire des bénéfices conduit à une priorisation des individus en fonction de leur revenu (*via* des pondérations en fonction du revenu ou des approches plus directement welfariste de type revenu équivalent monétisant l'ensemble des dimensions du bien-être). Les « valeurs sociales » ne rejoignent pas toujours ce principe de priorisation: pour prendre le cas de la santé, on peut trouver légitime un égal accès aux soins (indépendant du revenu) sans souhaiter une redistribution des revenus (ce qui, en règle générale, n'est pas compatible avec une approche welfariste de type revenu équivalent) ; on peut souhaiter valoriser également un gain en santé pour une personne jeune et une personne âgée (alors qu'une méthode de type revenu équivalent avec aversion aux inégalités priorisera l'individu jeune car la société a plus « à perdre » avec un individu jeune en mauvaise santé coûtant toute sa vie qu'avec une personne âgée). La fongibilité des dimensions du bien-être induite par le calcul coûts-bénéfices ne rencontre pas

¹⁷⁸ Kriström B. (2006), "Framework for assessing the distribution of financial effects of environmental policies", dans Serret Y. et Johnstone N. (éditeurs), *The Distributional Effects of Environmental Policy*, Cheltenham, Edward Elgar et OCDE, Paris.

¹⁷⁹ Manski C. F. (2015), "Narrow or broad cost-benefit analysis?", *Criminology & Public Policy* 14(4): 647-51.

¹⁸⁰ Fleurbaey M (2010), "Defense of the ex-post evaluation of risk", Note.

¹⁸¹ La perspective *ex post* pourrait être davantage compatible avec une approche visant une égalisation des réalisations effectives, tandis que la perspective *ex ante* le serait davantage avec une approche visant une égalisation des opportunités. Voir Fleurbaey M. (2008), *Fairness, Responsibility, and Welfare*. Oxford: Oxford University Press ; Fleurbaey M. et Peragine V. (2013), "Ex ante versus Ex post Equality of Opportunity", *Economica*, Vol. 80, p. 118-130.

nécessairement les « valeurs sociales » telles qu'elles peuvent s'exprimer dans des travaux cherchant à les révéler.

3.1. Répartition des effets entre les différentes parties prenantes

Une étape incontournable pour tenir compte des enjeux redistributifs consiste à répartir les coûts et les bénéfices entre des groupes de catégories socio-professionnelles ou de niveaux de revenu différents, entre les personnes vivant en milieu rural ou urbain, etc. Cette approche a l'intérêt de faciliter la prise de position du décideur, qui dispose ainsi d'une identification simple et claire des gagnants et des perdants. De façon plus importante, le décideur peut appliquer ensuite son propre système de pondérations et valoriser selon ses propres critères les populations qu'il souhaite.

Comme pour la plupart des éléments qui constituent une analyse coûts-bénéfices, ce type d'approche est conceptuellement très simple mais en pratique se révèle délicat. Comment comptabiliser les bénéfices et les coûts qui doivent être affectés à tel ou tel groupe en fonction de son niveau de revenu, etc. ? Une telle analyse présuppose que l'on dispose d'éléments très précis, voire de données individuelles.

Pour ce type d'analyses, l'utilisation de modèle de microsimulation est particulièrement adaptée afin d'affecter les effets d'une politique donnée pour différents groupes de personnes. C'est notamment ce qu'ont fait Kok *et al.* (2011)¹⁸² en simulant l'effet d'une politique de « désinvestissement social¹⁸³ » consistant à réduire de 650 millions d'euros les allocations familiales aux Pays-Bas. Le Tableau 3 ci-après montre les coûts et les bénéfices estimés par microsimulation du scénario envisagé (par rapport à la situation contrefactuelle prévalant). Les coûts et bénéfices sont répartis entre différents groupes. Kok *et al.* (2011) trouvent que les bénéfices perçus par certains peuvent correspondre à des coûts subis par d'autres. Par ailleurs, cette politique peut avoir un effet plus global *via* un « effet de second tour ». Certes, la politique a pour effet direct de réduire le recours aux modes de garde formels, mais cela affecte ensuite l'activité du secteur, induisant des suppressions d'emploi et des fermetures de structures d'accueil des jeunes enfants d'un même ordre de grandeur. Cette politique de réduction de la dépense publique de 650 millions d'euros pour l'accueil des jeunes enfants coûte en réalité 350 millions d'euros à la société¹⁸⁴.

Identifier avec suffisamment de précision les « gagnants » et les « perdants », ainsi que leurs revenus et/ou certains autres aspects de leur position relative au sein de la société, soulève en général des difficultés pratiques. Or sans cette information, aucune analyse générale des problèmes de répartition ne peut être envisagée. **C'est pourquoi les**

¹⁸² Kok L., Koopmans C., Berden C. et Dosker R. (2011), « De waarde van kinderopvang », Amsterdam: SEO Economisch Onderzoek; no. 2011-29).

¹⁸³ L'usage de ce terme un peu polémique est ici justifié puisque l'évaluation coûts-bénéfices montre que l'économie de dépense publique est plus que compensée par la perte sociale.

¹⁸⁴ Le modèle prédit la réaction des parents sur le marché du travail et un ajustement subséquent de la fiscalité. En résumé, sans subvention aux modes de garde, les parents se retirent partiellement du marché du travail, bénéficient de temps libre mais perdent en salaire et paient moins d'impôts. La hausse du coût des modes de garde a donc un effet partiellement positif sur l'économie, mais cet effet ne compense pas la perte de salaires.

considérations d'équité et de lutte contre les inégalités devraient être évoquées relativement tôt dans le processus de conception de la politique.

Tableau 1 – Microsimulation des effets d'une réduction des allocations familiales aux Pays-Bas pour un montant de 650 millions d'euros

	Enfant	Parents	Personnes en charge de la garde formelle	Personnes en charge de la garde informelle	Employeurs	État	Reste de la société	Total
Économies <i>ex ante</i>		-650				650		
Coût de la garde formelle			380					380
Coût de la garde informelle				-130				-130
Paiement de la garde formelle		380	-380					
Paiement de la garde informelle		-110		110				
Coût de la garde ludique		-5		20		-15		
Réduction des allocations pour mode de garde		-95				95		
Production					-740			-740
Salaires	?	-740			740			
Taxes		150				-150		
Coûts/Bénéfices		30				-30		
Taux d'imposition		175				-550	375	
Stimulant économique		45					95	140
Sous-total		-820					470	-350
Valeur du temps libre		380						380
Total	?	-440	0	0		0	470	30

Source : Kok et al. (2011)

3.2. Utiliser des pondérations pour intégrer l'équité dans l'analyse coûts-bénéfices

Comment intégrer dans l'analyse coûts-bénéfices les enjeux redistributifs ? Comment pondérer les bénéfices et les coûts sociaux imputables à différents groupes de personnes ? Ces pondérations peuvent en théorie être appliquées pour n'importe quel type de groupe, défini par exemple par son niveau de revenu, sa catégorie socio-professionnelle (ouvriers, employés, etc.), son lieu de résidence (rural ou urbain), etc. Alors que certaines caractéristiques individuelles sont facilement identifiables (telles que le genre, le niveau d'éducation, etc.), d'autres catégorisations laissent un plus grand degré de liberté à l'analyste : par exemple, les ménages peuvent être classifiés selon leur quintile de revenu. Dans ce cas, des analyses de sensibilités peuvent être menées pour voir si un tel choix n'a pas d'incidence sur les résultats du calcul.

Pour déterminer la valeur des pondérations, Loomis (2011)¹⁸⁵ suggère d'utiliser des décisions publiques ayant trait à des questions d'équité. Tel est le cas lorsqu'il s'agit de fixer des taux marginaux d'imposition. En ce cas, il convient de procéder à une normalisation. Supposons, par exemple, que pour les personnes gagnant moins de 10 000 euros, le taux marginal soit de 10 %, qu'entre 10 000 et 30 000 euros, il soit de 15 %, et au-dessus il soit égal à 30 %. En ce cas, cela revient à attribuer une pondération de 1,5 aux ménages imposés à un taux de 10 % et de 0,5 à ceux gagnant au-dessus de 30 000 euros.

Des travaux plus élaborés utilisent la théorie de la taxation optimale pour déduire les pondérations à partir des préférences révélées. Si cette approche peut paraître pertinente, elle n'est pas sans défaut. Elle peut notamment être biaisée du fait de l'existence d'électeurs médians, ce qui concrètement peut avoir pour conséquence des pondérations non monotones, voire négatives¹⁸⁶. Les électeurs médians ayant un rôle très important dans le résultat d'un vote, cela pourrait conduire à une forme de clientélisme en leur faveur, consistant par exemple à réduire leurs impôts. Les préférences révélées par la fiscalité refléteraient alors cette forme de clientélisme.

Les préférences exprimées, et notamment les consentements à payer ou à recevoir, peuvent également être utilisées pour établir les pondérations¹⁸⁷. L'étude de Samson *et al.* (2018)¹⁸⁸ présentée en annexe, propose une stratégie utilisant le revenu équivalent-santé. Loomis (2011) présente des exemples appliqués à des politiques environnementales.

Le CPB néerlandais utilise quatre groupes de pondérations issues de différentes méthodes¹⁸⁹. Leur étude montre notamment que la valeur estimée dans l'analyse coûts-bénéfices est sensible à la normalisation des pondérations. Par conséquent, les résultats déduits de normalisations différentes ne peuvent pas être comparés. Il existe toutefois des pondérations conduisant à un résultat pouvant être comparé avec le modèle sans pondérations. Il suffit pour cela d'utiliser une normalisation telle que :

$$\omega_i^n = \frac{1}{\sum_{j=1}^J P_j \omega_j} Y_i^{-\rho}$$

Dans cette expression, ω_i^n correspond à la pondération normalisée pour l'individu i , ω_j est la pondération avant normalisation pour le groupe de personnes j (il y a J groupes dans la population), P_j est la proportion de personnes dans le groupe j , Y_i est la valeur actuelle nette de la politique pour l'individu i et ρ est le paramètre d'aversion à l'inégalité. Si ce dernier vaut

¹⁸⁵ Loomis J. B. (2011), "Incorporating distribution into BCA using non-market valuation", *Journal of Costs-Benefits Analysis*, Vol. 2(1), p. 1-24.

¹⁸⁶ Zoutman F. T., Jacobs B. et Jongen E. L. W. (2016), "Redistributive politics and the tyranny of the middle class", CESifo Working Paper Series 5881, CESifo Group Munich.

¹⁸⁷ Loomis J. B., Bell P., Zita-Cooney H. et Asmus C. (2009), "A comparison of actual and hypothetical willingness to pay of parents and non-parents for protecting infant health: The case of nitrates in drinking water", *Journal of Agricultural and Applied Economics*, Vol. 41(3), p. 697-712.

¹⁸⁸ Samson A. L. *et al.* (2018), "Fairness in cost-benefit analysis: a methodology for health technology assessment", *Health Economics*, Vol. 27, p. 102-114.

¹⁸⁹ Van der Pol T., Bos F. et Romijn G. (2017), "Distributionally weighted cost-benefit analysis: From theory to practice", CPB Discussion Paper n° 364, Centraal Planbureau, Den Haag.

0, les pondérations valent toutes 1 et l'on retrouve la formule utilisée dans le cas de l'analyse coûts-bénéfices habituelle.

Que retenir ?

Les analyses coûts-bénéfices classiques ne tiennent pas compte de la redistribution. Pourtant, lorsque la VAN est positive, il est possible de redistribuer les gains. Toutefois de plus en plus de travaux intègrent les questions d'équité et de redistribution, même s'ils restent assez théoriques et éloignés de la décision politique.

La plupart de ces travaux proposent de pondérer les gains et les pertes de certains groupes au moyen d'une fonction de bien-être social fondée sur un objectif d'égalisation de la répartition de certains effets au sein de la population.

Les pratiques actuelles relatives aux enjeux d'équité dans l'analyse coûts-bénéfices relèvent le plus souvent d'une démarche ad hoc qui se limite le généralement à une présentation de la répartition des bénéfices entre différentes catégories de parties prenantes.

L'utilisation de modèles de microsimulation est particulièrement adaptée car elle permet de répartir les bénéfices et les coûts d'une politique donnée entre différents groupes de personnes.

Lorsque l'on construit des pondérations avec un objectif d'équité, la difficulté provient à la fois de la définition de ces pondérations et de la façon de les intégrer dans le calcul.

Pour les définir, les auteurs utilisent généralement des préférences révélées (dans la progressivité de la fiscalité par exemple) ou exprimées (à l'aide d'évaluations contingentes).

Pour les intégrer dans le calcul, il convient d'utiliser une fonction de bien-être social qui puisse être comparée avec une fonction sans pondérations, la normalisation des pondérations ayant dans la plupart des cas un effet sur la VAN obtenue.

4. Les effets d'équilibre général

Si certaines des politiques d'investissement social que nous avons considérées jusqu'ici peuvent être considérées comme des interventions limitées ou non systémiques, certaines autres peuvent avoir une incidence sur le reste de l'économie par le biais de ce que l'on appelle parfois des effets de second tour. Dans ce cas, pour être complet, un calcul socioéconomique devrait tenir compte des effets « d'équilibre général » de la politique envisagée. Bien souvent, les évaluations de politiques publiques estiment et sélectionnent un sous-ensemble d'effets de ces politiques, principalement ceux qui concernent l'emploi, la santé, les conditions de vie, etc., des bénéficiaires de ces politiques en les comparant à la situation des non-bénéficiaires comparables : ces évaluations correspondent à ce que l'on nomme couramment une analyse en équilibre partiel. Des analyses en équilibre général devraient prendre en compte la plupart ou tous les effets de ces politiques, notamment les effets qu'elles pourraient avoir, à court ou à plus long terme, sur l'ensemble de la société ou

de l'économie. En pratique, le cadre utilisé par la plupart des analyses se situe entre un cadre d'équilibre partiel assez strict, qui examine uniquement le groupe ciblé, et un modèle d'équilibre général entièrement calculable, dont la modélisation demeure sujette à caution, compte-tenu des nombreuses hypothèses qui président à son élaboration.

4.1. Évaluer un modèle intermédiaire, entre équilibre général et équilibre partiel

Le terme « équilibre général » peut en pratique correspondre à plusieurs types d'analyses. Une analyse en équilibre général peut prendre en compte les effets de la politique évaluée sur plusieurs marchés ou décrire uniquement les effets de cette politique sur les prix, les revenus et les services non marchands affectés par cette politique.

Une analyse élargie peut également prendre en compte les effets d'une intervention ciblée sur les groupes de personnes non directement bénéficiaires. C'est le cas par exemple des politiques d'accompagnement et d'aide à la recherche d'emploi destinées à un groupe particulier de chômeurs. Faciliter le retour à l'emploi de certains peut se faire au détriment d'autres : face à une demande de travail dont le niveau ne varie pas dans le court terme, une intervention de ce type peut certes améliorer les chances de réemploi rapide des personnes ciblées (« traitées »), mais son effet global n'est probablement qu'une réaffectation des rangs dans la file d'attente de sortie du chômage, les chômeurs non bénéficiaires de l'intervention, devenus « non prioritaires », perdant des places dans cette file. Ce type de phénomène a été mis en évidence par Crépon *et al.* (2013), qui, comme cela a été rappelé plus haut, ont étudié les effets d'un programme réservé aux jeunes chômeurs diplômés. Ce programme s'est certes révélé efficace à court terme pour ce groupe de chômeurs. Mais cet effet positif s'est fait au détriment des chômeurs non bénéficiaires, plus âgés ou moins diplômés, porteurs d'une plus grande expérience professionnelle ou titulaires d'un niveau de diplôme plus élevé, et donc potentiellement substituables aux jeunes chômeurs diplômés. En d'autres termes, dans les bassins d'emploi où le nombre de créations d'emploi est faible, plus la proportion de jeunes chômeurs bénéficiaires augmente, plus ceux qui sont aléatoirement écartés du programme trouvent difficilement un emploi.

La plupart des évaluations de politique sociale utilisent un cadre d'équilibre partiel modifié qui prend en compte les effets de cette politique sur plusieurs marchés supposés être indépendants. Une autre solution consiste à modéliser les interdépendances au sein de plusieurs marchés sélectionnés par l'analyste (tel le marché du travail), ou entre ces marchés (par exemple, entre le marché de l'emploi et le marché du logement), ce qui constitue un compromis entre les limitations pratiques de données ou de calcul, et la nécessité d'être exhaustif (dans l'optique d'une analyse idéale d'équilibre général).

Cette approche nécessite beaucoup moins de ressources que les modèles complets d'équilibre général et peut fournir une aide appropriée pour une prise de décision rationnelle. Dans de nombreux cas, les enjeux ne justifient pas la construction d'un modèle complet d'équilibre général. De petites interventions fondées sur des quasi-expériences ou des recherches de terrain peuvent être modélisées de manière efficace et avec une précision raisonnable sans la complexité d'un cadre d'équilibre général.

La prise en compte d'effets d'équilibre général peut avoir d'importantes conséquences sur les résultats du calcul socioéconomique, notamment lorsque des distorsions ou des amenuisements des effets de la politique évaluée apparaissent à la suite de leurs conséquences sur certains marchés (voir Goulder et Williams, 2003, Fullerton et Heutel, 2011, ou Klaiber et Smith, 2010)¹⁹⁰. Une politique qui est estimée bénéfique lorsqu'elle est évaluée sur un marché spécifique, peut se révéler inefficace lorsque l'analyse est étendue à d'autres marchés.

4.2. Quels marchés retenir ?

Un modèle complet d'équilibre général est par définition une construction théorique. S'il est toujours possible de conduire une simulation permettant de déduire à l'aide d'un tel modèle les effets de la politique évaluée, l'exercice reste largement formel, soumis à de nombreuses hypothèses comportementales et fonctionnelles. L'objectif d'une analyse d'équilibre général devrait de ce fait être plus modeste : plutôt que d'essayer de prévoir, en l'absence d'une inférence causale difficile à conduire, les effets d'une politique sociale sur l'ensemble des marchés, il conviendrait de limiter l'analyse de ses effets à un nombre limité de sphères socioéconomiques, connexes à celles susceptibles d'être directement impactées par la politique considérée. À ce propos, Goulder et Williams (2003) soulignent que l'importance de la prise en compte des effets d'équilibre général est largement indépendante de la taille du marché ciblé par la politique (par exemple, le marché du travail). Les facteurs justifiant d'étendre l'analyse à d'autres marchés qu'au marché ciblé sont les suivants : (1) sur ces marchés connexes (santé, logement, délinquance, etc.), le bénéfice social marginal et le coût social marginal de la politique considérée doivent être significativement différents ; des études préalables peuvent permettre d'avoir une idée de la validité de cette hypothèse ; (2) la dimension du marché affecté doit être suffisante; cela exclut des marchés très locaux (par exemple, le marché de l'immobilier dans certaines zones peu denses) ; et (3) ce marché doit entretenir un lien étroit avec le marché ciblé (en d'autres termes, l'élasticité croisée des comportements sur ces deux types de marchés doit avoir été estimée comme significative dans des études antérieures). En conséquence, le recours à une analyse des effets d'équilibre général doit être fonction : (1) de la disponibilité des données et des ressources requises pour l'identification de ces effets ; et (2) de la présence de marchés connexes significativement affectés (voir Goulder et Williams, 2003).

Afin de présenter un degré de crédibilité satisfaisant, l'analyse devrait être raisonnablement proportionnée, à mi-chemin entre l'équilibre trop partiel et le modèle d'équilibre général par trop hypothétique. Le calcul socioéconomique ne peut prendre en compte tous les effets d'équilibre général. Priorité devrait être donnée à ceux des effets d'équilibre général qui sont en lien direct avec les conséquences attendues de la politique évaluée.

¹⁹⁰ Goulder L. et Williams R. (2003), "The substantial bias from ignoring general equilibrium effects in estimating excess burden, and a practical solution", *Journal of Political Economy*, Vol. 111(4), p. 898-927 ; Fullerton D. et Heutel G. (2011), "Analytical general equilibrium effects of energy policy on output and factor prices", *The B.E. Journal of Economic Analysis et Policy*, Vol. 10(2), p. 1-26 ; Klaiber H. A. et Smith V.K. (2010), "General equilibrium benefit analyses for social programs", dans *Towards Principles and Standards in the Use of Benefit-Cost Analysis: A Summary of Work, and a Starting Place*, édité par Zerby R. et al., Seattle, WA: Benefit-Cost Analysis Center, University of Washington.

4.3. L'exemple des bourses d'enseignement supérieur

Prenons l'exemple d'une mesure visant à accroître le nombre d'étudiants dans les universités par le biais d'une augmentation du nombre de bourses d'études offertes. Nous avons dans le chapitre 3 évoqué le cas de la France, avec l'article de Fack et Grenet (2015)¹⁹¹. Afin de présenter les résultats d'un modèle d'équilibre général tel que celui développé par Heckman *et al.* (1998)¹⁹², calibré à partir de statistiques provenant des États-Unis, il semble préférable de s'appuyer sur l'article pionnier de Dynarski (2004)¹⁹³, consacré au programme Hope ("*Helping Outstanding Pupils Educationally*") introduit en 1993 en Géorgie. Ce programme offrait des bourses aux étudiants s'inscrivant en première année de *college* (correspondant à la 1^{re} année de licence en France) et satisfaisant aux conditions d'éligibilité suivantes : 1) ils ou elles devaient être diplômés d'une *high school* (lycée) de Géorgie avec un GPA (*Grade Point Average*) d'au moins 3 sur 4, et ils ou elles devaient s'inscrire dans un *college* de cet État ; 2) ils ou elles devaient obtenir ce même niveau de GPA au *college* pour continuer à percevoir la bourse ; et 3) le revenu de leurs parents devait être inférieur à 66 000 dollars en 1993 (à 100 000 dollars en 1995). Le montant de la bourse était de 500 dollars en 1993, de 3 000 dollars en 1996, plus un voucher de 500 dollars pour ceux ayant obtenu le *General Education Diploma* (GED). À l'aide de la méthode des doubles différences, Dynarski (2004) a comparé les proportions de jeunes de 18 et 19 ans inscrits en 1^{re} année de *college* en Géorgie et dans les autres États du Sud, avant et après mise en place du programme. Elle a trouvé que la bourse HOPE a eu pour effet de faire croître le taux d'inscription en 1^{re} année de *college* de 5 à 7 points de pourcentage.

Que nous enseigne un modèle économique d'équilibre général dans lequel on simule une baisse substantielle et généralisée des frais de scolarité ? Ses prédictions confirment-elles les résultats des évaluations faites avec des données d'enquête, telles que celles obtenues par Dynarski (2004) ? Le modèle d'équilibre général d'Heckman *et al.* (1998) est un modèle dynamique qui permet de quantifier les effets sur les salaires de deux catégories de salariés, les diplômés et les non-diplômés. Une réduction de 500 dollars des frais de scolarité est financée par une hausse proportionnelle de l'impôt sur le revenu¹⁹⁴. Cette augmentation de l'impôt peut réduire les incitations à faire des études supérieures. En équilibre partiel, lorsque les salaires relatifs des deux catégories de travailleurs sont maintenus constants, Heckman *et al.* (1998) ont trouvé que la réduction des frais d'inscription induisait une hausse de 5,3 % des inscriptions à l'université, valeur comparable à celle estimée par Dynarski (2004). Toutefois, dans le cadre du modèle d'équilibre général, cette hausse des inscriptions n'est que de 0,46 % parce que le salaire relatif des diplômés baisse significativement.

¹⁹¹ Fack G. et Grenet J. (2015), "Improving College Access and Success for Low-Income Students: Evidence from a Large Need-Based Grant Program", *American Economic Journal: Applied Economics*, Vol.7(2), p. 1-34.

¹⁹² Heckman J. J., Lochner L. et Taber C. (1998), "General-equilibrium treatment effects: A study of tuition policy", *American Economic Review*, Vol. 88(2), p.381-386.

¹⁹³ Dynarski S. (2004), "The New Merit Aid", dans *College Choices: The Economics of Where to Go, When to Go, and How to Pay for It*, édité par C. Hoxby, University of Chicago Press.

¹⁹⁴ Cette hypothèse simplifie grandement le modèle.

4.4. Prendre en compte les externalités : le cas des bénéfices sociaux de l'éducation

Les modèles en équilibre général reposent le plus souvent sur des hypothèses fortes, relatives aux comportements des agents ou aux formes mathématiques des fonctions d'utilité ou de profit utilisées pour la construction des modèles. Ces hypothèses sont difficilement vérifiables, voire pour certaines d'entre elles peu réalistes. Cette critique souvent répétée incite les économistes à mettre en évidence des effets moins généraux, qui vont toutefois au-delà des effets du dispositif ou de la politique évaluée sur ses seuls bénéficiaires. L'analyse concerne alors les externalités, positives ou négatives, de la politique considérée. L'exemple de l'éducation est de ce point de vue particulièrement instructif.

Les économistes ont longtemps débattu pour savoir si l'éducation est un bien privé, qui profite exclusivement à celui ou celle qui investit en ce domaine, ou si elle peut produire des externalités positives, se diffuser au sein de la société, notamment dans les lieux de travail ou dans les familles, et de ce fait être un bien public qui profite à ceux qui sont moins éduqués¹⁹⁵. Les travaux empiriques de Moretti (2004) et de Mas et Moretti (2009) ont montré que cette dernière hypothèse semble bien vérifiée¹⁹⁶. Moretti (2004) a trouvé que les salaires des personnes ayant un niveau d'éducation secondaire (lycée) sont supérieurs dans les zones d'emploi dans lesquelles la proportion de salariés ayant un niveau d'éducation supérieur (post-secondaire) est plus élevée. S'agit-il d'une simple corrélation, provenant d'un troisième facteur, par exemple du fait que la zone d'emploi abrite des activités particulièrement productives ou d'un contexte politico-économique favorable aux salariés, ou d'une véritable relation de causalité ? Le modèle longitudinal de Moretti (2004) tient compte des effets de sélection qui pourraient produire une concentration géographique des plus éduqués dans certaines zones d'emploi. Pour ce faire, il considère deux variables instrumentales qui affectent la proportion de salariés très éduqués sans pour autant agir directement sur le niveau des salaires des moins éduqués : ces deux instruments sont la structure démographique passée et la présence d'une université de type *Land-Grant*¹⁹⁷ dans la localité. Moretti (2004) a trouvé qu'une augmentation d'un point de pourcentage de salariés titulaires d'un diplôme de *Bachelor* (licence) accroît de 1,9 % les salaires des travailleurs ayant quitté le lycée avant la fin des études secondaires, de 1,6 % le salaire de ceux ayant été au terme des études secondaires, et de 0,4 % le salaire de ceux ayant fait des études supérieures : les moins éduqués sont donc les principaux gagnants de l'accroissement de la proportion de salariés ayant fait des études supérieures. Des données appariées entreprises-salariés permettent d'affirmer que ces accroissements de salaires proviennent des gains de productivité engendrés par l'emploi de travailleurs plus éduqués. Mas et Moretti (2009) ont montré empiriquement que ces gains de productivité sont tout à la

¹⁹⁵ La communication de Robert Gary-Bobo devant le groupe de travail faisait référence à ce débat.

¹⁹⁶ Moretti E. (2004), "Estimating the social return to higher education", *Journal of Econometrics*, Vol. 121(1-2), p. 175-212 ; Mas A. et Moretti E. (2009), "Peers at work", *American Economic Review*, Vol. 99(1), p. 112-145.

¹⁹⁷ Les *Land-Grant colleges* sont des institutions américaines d'enseignement supérieur, désignées par chaque État pour recevoir le bénéfice du *Morrill Land-Grant Colleges Act*. Les *Morrill Land-Grant Colleges Acts* sont une série de lois américaines ayant institué des dons de terrains fédéraux à chaque État dans le but de mettre en place des institutions d'enseignement supérieur. La plupart de ces *land-grant colleges* ont été créés au XIX^e siècle ou au début du XX^e siècle.

fois le produit de la pression qu'exercent au sein de l'entreprise les cadres plus éduqués¹⁹⁸ sur les employés moins éduqués, mais aussi de la diffusion des savoirs que les premiers dispensent aux seconds. Les travaux récents sur les effets de pairs au sein des collectifs de travail confirment ces résultats¹⁹⁹.

Les gouvernements investissent beaucoup en éducation et la question est de savoir si, du point de vue de l'État, cet investissement est rentable. Certains chercheurs ont proposé des analyses coûts-bénéfices de ces investissements en tenant seulement compte des variations de dépenses et recettes socio-fiscales et judiciaires par exemple dans le cas du Texas et de la Californie. Murdock *et al.* (2003)²⁰⁰ ont calculé le total des dépenses publiques engagées pour l'enseignement supérieur au Texas. Ils ont estimé que les effets combinés d'une éducation supérieure sur un moindre recours à l'aide sociale, une moindre propension à la délinquance et une plus grande contribution aux prestations sociales et à l'impôt sur le revenu *via* des salaires plus élevés, correspondait à un rapport entre bénéfices et coûts égal à 4 : 1 dollar supplémentaire investi dans l'enseignement supérieur rapporte 4 dollars à terme. Brady *et al.* (2005)²⁰¹ ont fait le même calcul pour la Californie : ils trouvent que le rendement net est de 3,65 dollars pour un dollar dépensé.

4.5. Équilibre partiel, équilibre général : l'exemple d'une politique éducative

À titre d'illustration, Klaiber et Smith (2012)²⁰² ont pris l'exemple des effets de la qualité des écoles sur les choix de résidence des familles, et donc sur le marché du logement, dans le comté de Maricopa (Arizona), dont le siège est Phoenix. Ce comté, comprenant 46 secteurs scolaires (*school districts*), est le plus peuplé d'Arizona. En 2005, sa population était de 3,6 millions d'habitants, ce qui en fait le quatrième comté des États-Unis. La période retenue couvre les années 2003 à 2006.

Les données utilisées

Les données utilisées contiennent des informations sur l'ensemble des transactions immobilières réalisées par des particuliers, mais aussi sur la qualité des établissements scolaires, mesurée à partir des rapports annuels fournis par ces établissements. Les rapports scolaires annuels contiennent, pour chaque établissement, le nombre d'enseignants et d'aides-enseignants, le nombre d'élèves scolarisés dans l'établissement ainsi que leurs résultats à des tests standardisés en mathématiques, en anglais et en lecture. L'analyse prend en compte les résultats des tests pour les grades 2 (correspondant au CE1 en France) à 8 (correspondant à la classe de troisième en France) et se limite aux scores des élèves anglophones.

¹⁹⁸ Cette analyse ignore les cadres moins éduqués ayant bénéficié de promotions internes.

¹⁹⁹ Voir, par exemple, Bandiera O., Barankay I. et Rasul I. (2009), "Social Connections and Incentives in the Workplace: Evidence from Personnel Data", *Econometrica*, Vol. 77(4), p. 1047-94.

²⁰⁰ Murdock S., White S., Hoque N., Pecotte B., You X. et Balkan J. (2003), *The New Texas Challenge: Population Change and the Future of Texas*. College Station: Texas A&M University Press.

²⁰¹ Brady H. B., Hout M. et Stiles J. (2005), *Return on Investment: Educational Choices and Demographic Change in California's Future*. Survey Research Center, University of California Berkeley.

²⁰² Klaiber H.A. et Smith V.K. (2012), "Developing General Equilibrium Benefit Analyses for Social Programs: An Introduction and Example", *Journal of Benefit-Cost Analysis*, Vol. 3(2), p. 1-50.

Klaiber et Smith (2012) ont évalué les nombres moyens d'élèves, d'enseignants et d'aides-enseignants pour chaque combinaison classe/secteur scolaire/année ainsi que la moyenne des scores en mathématiques, en lecture et en anglais pour chaque combinaison année/établissement scolaire/année. Les données individuelles sur les établissements scolaires ont été ensuite agrégées au niveau du secteur scolaire, selon le type de test, du niveau de la classe et de l'année. En utilisant ces moyennes, Klaiber et Smith (2012) ont estimé le nombre moyen d'élèves par enseignant ainsi que le nombre moyen d'élèves par aide-enseignant. Leur échantillon se compose au total de 3 711 combinaisons niveau de la classe/secteur scolaire/année/type de test.

La mesure de la qualité de l'école utilisée dans l'estimation du modèle théorique, qui est un modèle d'assortiment (*sorting model*) est déduite d'un modèle économétrique expliquant les résultats des élèves aux tests par le type de test, le niveau d'enseignement, l'année, et par les nombres moyens d'élèves par enseignant et par aide-enseignant dans l'établissement scolaire considéré²⁰³. Ce modèle permet de prévoir l'effet d'une réduction du nombre d'enseignants sur la qualité des résultats scolaires, établissement par établissement. À cet égard, Klaiber et Smith (2012) ont retenu comme mesure de la qualité de l'établissement scolaire la valeur prédite du résultat au test d'anglais pour les élèves du grade 2 (CE1) en 2003.

Un modèle log-linéaire a été utilisé pour estimer un indice de prix du logement au sein de chaque secteur scolaire. Parmi les variables explicatives retenues figurent les caractéristiques de chaque logement vendu, l'année de la transaction et des effets fixes représentant les secteurs scolaires. L'échantillon utilisé contenait 406 556 transactions immobilières enregistrées entre 2003 et 2006. Il existe une forte corrélation positive entre les variations des indices représentant la qualité des établissements scolaires et les variations de ceux mesurant le prix des logements dans les 46 secteurs scolaires.

La population et le niveau des revenus dans chaque secteur scolaire ont été obtenus grâce au recensement de l'année 2000. Les revenus ont été classés en 16 catégories de niveaux croissants, les revenus les plus bas étant ceux inférieurs à 10 000 \$, les plus élevés étant ceux supérieurs à 200 000 \$.

Le modèle

Le modèle théorique sous-jacent est un modèle d'assortiment (*sorting model*). Chaque ménage est supposé choisir un logement dans l'un des secteurs scolaires caractérisé par des établissements scolaires de qualité connue.

La décision du ménage est prise en deux temps. Celui-ci évalue tout d'abord ce que serait son choix dans chaque secteur en fonction du prix des logements et des biens publics locaux dans ce secteur. Ensuite, le ménage évalue l'utilité associée à chaque choix et décide de se loger dans le secteur qui lui procure la plus grande utilité, selon son niveau de revenu²⁰⁴. Le modèle prédit la proportion de ménages qui décide d'acquérir un logement

²⁰³ Le modèle ne prend pas en compte la composition sociale des établissements scolaires. C'est là une de ses limites.

²⁰⁴ La fonction d'utilité du ménage est une fonction avec élasticités de substitution constantes (de type CES généralisée). Le modèle est estimé à l'aide de la méthode des moments généralisés (GMM) : les moments

dans chaque secteur scolaire. Les estimations de Klaiber et Smith (2012) sont comparables à celles précédemment obtenues, notamment par Bayer *et al.* (2007)²⁰⁵ : la disponibilité marginale à payer pour un logement situé dans un secteur scolaire dont la qualité est supérieure d'un écart-type est comprise entre 5,24 et 10,86 \$ par mois (en dollars de 1990).

Les effets simulés d'une réduction du nombre d'enseignants

En avril 2009, plus de 1 600 enseignants ont été licenciés dans le comté de Maricopa (Arizona). Ces réductions d'effectifs ont été inégalement réparties dans les secteurs scolaires du comté. Elles ont eu pour effet direct d'augmenter le nombre d'élèves par enseignant et donc en théorie de dégrader les résultats scolaires dans les établissements concernés. Si les coûts de mobilité résidentielle sont faibles, la réduction du nombre d'enseignants incite les familles des élèves concernés à déménager pour aller habiter dans des secteurs scolaires non touchés par cette mesure. Ces mobilités devraient avoir pour conséquence de faire augmenter les prix de l'immobilier dans les secteurs mieux dotés en enseignants, mais aussi de faire baisser ces prix dans les secteurs dans lesquels le nombre d'enseignants a diminué. En raison de cette variation relative des prix des logements, la propension à déménager devrait alors être freinée dans les secteurs scolaires où le nombre d'enseignants a été initialement réduit.

L'approche en équilibre partiel rend seulement compte de la réduction de la qualité de l'enseignement dans les secteurs scolaires initialement affectés. À l'équilibre général, ces différences s'estompent car, dans les secteurs où le nombre d'enseignants n'a pas été réduit, le nombre moyen d'élèves par enseignant a tendance à augmenter en raison de l'afflux des familles ayant quitté les secteurs scolaires moins dotés. À terme, en raison de l'augmentation du nombre moyen d'élèves par enseignant dans ces secteurs, le niveau des résultats scolaires devrait également y être altéré. Au total, en raison des effets de second tour²⁰⁶, la qualité de l'enseignement devrait être moindre dans tous les secteurs scolaires, qu'ils aient été concernés initialement ou non par la réduction du nombre d'enseignants. Ses résultats dépendent cruciallement d'au moins deux hypothèses fortes, la faiblesse des coûts de mobilité résidentielle et l'absence d'effets de pairs et de ségrégation scolaire au sein des écoles. Si ces deux hypothèses étaient relâchées, les résultats de l'analyse risqueraient fort d'en être modifiés.

Pour être complet, un calcul socioéconomique devrait tenir compte des effets d'équilibre général de la politique envisagée.

Toutefois, il ne peut prendre en compte tous les effets d'équilibre général. En pratique, le cadre utilisé devrait se situer entre un cadre d'équilibre partiel strict, qui examine uniquement le groupe ciblé, et un modèle d'équilibre général entièrement calculable, dont les hypothèses demeurent souvent sujettes à caution.

considérés sont les trois quartiles des distributions des revenus et des dépenses consacrés à l'achat d'un logement.

²⁰⁵ Bayer P., Ferreira F. et McMillan R (2007), "A unified framework for measuring preferences for schools and neighborhoods", *Journal of Political Economy*, Vol. 115(4), p. 588-638.

²⁰⁶ Dans les faits, tout mécanisme relativement complexe ayant des conséquences à moyen terme peut être considéré comme engendrant un ensemble d'effets de second tour.

Cette approche nécessite beaucoup moins de ressources que les modèles complets d'équilibre général et peut fournir une aide appropriée pour une prise de décision rationnelle.

Priorité devrait donc être donnée à ceux des effets d'équilibre général qui sont en lien direct avec les conséquences attendues de la politique évaluée. En d'autres termes, le recours à une analyse des effets d'équilibre général doit être fonction : (1) de la disponibilité des données et des ressources requises pour l'identification de ces effets, et (2) de la présence de marchés connexes significativement affectés.

Que retenir ?

Pour être complet, un calcul socioéconomique devrait tenir compte des effets d'équilibre général de la politique envisagée.

Toutefois, il ne peut prendre en compte tous les effets d'équilibre général. En pratique, le cadre utilisé devrait se situer entre un cadre d'équilibre partiel strict, qui examine uniquement le groupe ciblé, et un modèle d'équilibre général entièrement calculable, dont les hypothèses demeurent souvent sujettes à caution.

Cette approche nécessite beaucoup moins de ressources que les modèles complets d'équilibre général et peut fournir une aide appropriée pour une prise de décision rationnelle.

Priorité devrait donc être donnée à ceux des effets d'équilibre général qui sont en lien direct avec les conséquences attendues de la politique évaluée. En d'autres termes, le recours à une analyse des effets d'équilibre général doit être fonction : (1) de la disponibilité des données et des ressources requises pour l'identification de ces effets, et (2) de la présence de marchés connexes significativement affectés.

Conclusions et recommandations

Appliquer le calcul socioéconomique aux politiques d'investissement social, telles que les politiques éducatives, les politiques d'emploi ou de santé, est encore chose peu courante, notamment en France où le retard pris en ce domaine par rapport à d'autres pays, nord-américains ou européens, est conséquent. La raison première en est que l'évaluation des effets des politiques publiques est dans notre pays trop peu fréquente, peu systématique, même si l'on constate une évolution récente en sa faveur²⁰⁷. **Or le calcul socioéconomique, qu'il se décline sous la forme de l'analyse coûts-bénéfices ou du ratio coûts-efficacité, est étroitement lié à la pratique de l'évaluation, notamment des méthodes d'évaluation d'impact. Il doit reposer, et c'est là un préalable devenu indispensable, sur des estimations robustes des effets à court et moyen termes de la politique considérée, celles-ci pouvant être obtenues par des évaluations expérimentales ou quasi-expérimentales, et, si ces observations directes ne sont pas suffisantes, par des techniques de microsimulation. Les revues extensives de littérature et les méta-analyses qui font la synthèse des résultats obtenus à l'échelle internationale peuvent également fournir des ordres de grandeur pour les paramètres du calcul qui ne peuvent être estimés à partir d'échantillons ou de fichiers de données françaises, administratives ou non.** Mais le calcul socioéconomique va au-delà de l'évaluation d'impact. S'il s'apparente par de nombreux aspects à l'analyse coûts-bénéfices appliquée aux investissements publics en matière d'infrastructure, de transport, d'énergie, etc., si son formalisme est souvent similaire, il s'en distingue de plusieurs manières.

Les spécificités du calcul socioéconomique appliqué à l'investissement social

1. Premièrement, les politiques d'investissement social concernent avant tout des personnes, des étudiants, des salariés, des ménages. Cela signifie que leurs effets sont à rechercher dans de multiples dimensions, toutes reliées à l'activité et aux conditions de vie des individus, qu'ils soient ou non bénéficiaires de la politique évaluée. Certes, les politiques d'investissement social comme les politiques d'investissement public, ont des effets tout à la fois de court et de plus long terme, des effets directs et plus indirects, des effets aisément valorisables, par exemple sur les salaires, sur les impôts et contributions sociales, sur les dépenses de santé, etc. Mais les politiques d'investissement social ont des effets plus spécifiques et qui sont moins facilement valorisables, tels que les effets sur la propension à la délinquance, sur la maltraitance des proches, sur le plus grand

²⁰⁷ Bono P-H., Debu S., Desplatz R., Hayet M., Lacouette-Fougère C. et Trannoy A. (2018), « Vingt ans d'évaluations d'impact en France et à l'étranger. Analyse quantitative de la production scientifique », Document de travail, N°2018-06, décembre, France Stratégie.

soin apporté à l'éducation des enfants, sur une plus grande participation à la vie citoyenne, etc. Ces effets de natures diverses sont souvent observés dans le long terme, par exemple à l'âge adulte lorsque l'on s'intéresse aux effets des politiques d'investissement social dans la petite enfance, comme l'ont montré les études internationales consacrées à ce thème. C'est pour cela que le suivi de long terme des effets des politiques d'investissement social est particulièrement nécessaire. Les effets induits qui ne peuvent être quantifiés doivent néanmoins être décrits de façon qualitative.

2. Une deuxième caractéristique, souvent passée sous silence, distingue le calcul socioéconomique appliqué aux politiques d'investissement social. Dans le cas de politiques d'investissement social, le calcul socioéconomique permet un pilotage plus fin des politiques concernées, qu'il s'agisse de politiques éducatives, de politiques d'emploi ou de santé. L'accumulation, en temps quasi continu, des mesures des effets d'une de ces politiques permet de la reconfigurer ou de la déployer différemment dans un bref délai, ce qui est évidemment plus difficile à réaliser dans le cas d'un investissement public fixe (une école, un hôpital, une route, une prison, etc.).
3. Une troisième différence doit être soulignée : plus encore que dans le cas des grands investissements économiques, les politiques sociales ont des effets variables, hétérogènes, pour des groupes de personnes différents²⁰⁸. La recherche internationale en sciences sociales a illustré ce fait par de nombreux exemples. Une politique éducative (par exemple, la réduction de la taille des classes ou l'enseignement par le numérique) aura des effets différenciés sur les enfants de familles défavorisées, sur les filles, sur les élèves les plus jeunes, etc. Il en est de même pour les politiques d'emploi et de prévention en santé. La conception et les modalités de la politique envisagée, le principe de son évaluation même, doivent en tenir compte. Une analyse globale, livrant un chiffre moyen, peut masquer des disparités assez fortes entre groupes d'individus. Mesurer cette hétérogénéité des effets devrait permettre de réviser les conditions d'éligibilité à la politique considérée et mieux cibler celle-ci. Cette opération peut toutefois être faite plus en amont, les résultats des évaluations menées dans d'autres pays sur des politiques similaires pouvant ici servir de guide.

Il faut donc éviter deux écueils, qui, remarquons-le, ne sont pas spécifiques à l'investissement social (même si le deuxième est sans doute plus prégnant) : premièrement, livrer un chiffre unique qui ne retiendrait qu'une seule valeur pour le ratio moyen entre bénéfices et coûts d'une politique publique, ce calcul pouvant dépendre des valeurs possibles d'un paramètre affectant les bénéfices futurs (taux d'actualisation, taux de progression des salaires, du chômage, etc.) ; deuxièmement négliger la dispersion de cette valeur moyenne, elle-même variable, entre groupes d'individus différents, alors même que coexistent des enjeux d'équité et de ciblage des politiques dans un souci d'efficience.

Disposer de données longitudinales pertinentes

Appliquer de manière crédible et rigoureuse le calcul socioéconomique aux politiques d'investissement social impose de disposer d'informations statistiques longitudinales provenant de différentes sources et susceptibles d'être appariées, de manière exacte en

²⁰⁸ Ce point a été particulièrement souligné par Cyril Nouveau dans sa communication au groupe de travail.

suivant *les mêmes individus*, ou moins exacte en utilisant, à partir de fichiers administratifs ou de sources statistiques publiques, les observations relatives à *des individus comparables* à ceux directement concernés par la politique évaluée. Pour le dire autrement, le calcul socioéconomique appliqué aux politiques d'investissement social est très exigeant en matière de données individuelles, l'évaluation des politiques d'investissement public pouvant se contenter plus facilement de données agrégées. De ce point de vue, il est possible d'affirmer que le système statistique public français, en raison de sa qualité, fournit déjà une masse considérable d'informations utiles, qu'il s'agisse des enquêtes réalisées par l'Insee, les services statistiques ministériels, l'Ined, l'Inserm, etc., et des fichiers administratifs constitués par les différents services statistiques des ministères et organismes publics (Cnaf, Pôle Emploi, etc.). Le principal problème est ici bien souvent qu'il est difficile aux chercheurs d'obtenir des appariements des informations contenues dans ces différentes sources de données, sans parler de l'accès direct à ces données qui, malgré de grandes avancées, reste malaisé dans certains domaines tels que la santé, la délinquance ou la justice. L'anonymisation des identifiants est pourtant possible, les techniques et la législation en ce domaine ayant significativement progressé. Le rapport sur l'accès des données administratives aux chercheurs, rédigé par Bozio et Geoffard (2017)²⁰⁹, fait le constat de l'absence d'obstacles juridiques – avec quelques exceptions – à l'accès aux données par les chercheurs. Il constate toutefois que les difficultés pratiques restent notables, et que le droit relatif à l'accès aux données administratives à des fins de recherche est certes complexe, mais surtout très mal connu des administrations productrices comme des chercheurs. Entre autres choses, le rapport de Bozio et Geoffard (2017) recommande la réalisation d'un guide des dispositifs juridiques en vigueur, et il préconise la création d'une instance de concertation impliquant les producteurs de données, les chercheurs, et les administrations compétentes pour la protection et l'accès aux données publiques.

Expérimenter pour mieux évaluer les effets de court terme

Au problème de l'accès aux données longitudinales, s'ajoute celui de la connaissance immédiate des effets de court terme d'une politique d'investissement social, connaissance qui peut, et même devrait, provenir de la mise en place plus systématique d'expérimentations randomisées, au niveau national ou local. De telles expérimentations permettent en effet d'appréhender sous l'angle de la causalité, et non plus de la simple corrélation, les bénéfices immédiats d'une politique de ce type. Ces expérimentations pourraient être plus souvent précédées d'essais-pilotes qui ont pour vertu d'être plus facilement gérables et qui livrent des conclusions permettant de mieux définir le pourtour, le contenu et le protocole des expérimentations à mettre en œuvre sur une plus grande échelle. Mais le temps de mise en place et d'exploitation des résultats de telles expérimentations, et de leurs éventuels pilotes, souvent compris entre deux et trois ans, est supérieur à celui du cycle politique et de ses exigences. Tenter de concilier au mieux ces deux temporalités, celle de la décision politique et celle de l'analyse scientifique, est certes

²⁰⁹ Bozio A. et Geoffard P. Y. (2017), *L'accès des chercheurs aux données administratives. État des lieux et propositions d'actions*, Rapport pour le Secrétariat d'État à l'industrie, au numérique et à l'innovation, Conseil national de l'information statistique, La Documentation française, Paris, 68 pages.

difficile, mais cet effort constituerait un progrès dans la gestion de la gouvernance de la politique publique.

Si, au terme d'une expérimentation randomisée, dont la réalisation et l'analyse des résultats s'étalent généralement sur trois ou quatre années, une politique d'investissement social ne produit pas les effets attendus, elle devrait être pour le moins réformée, voire abandonnée. Ainsi, de nombreuses politiques actives d'emploi, introduites dans plusieurs pays durant ces dernières décennies pour lutter contre le chômage de masse ou de longue durée, se sont révélées coûteuses et inefficaces au terme d'une analyse expérimentale ou quasi-expérimentale. La rationalité économique eût voulu qu'elles ne fussent pas prolongées.

Externalités et investissement social

La question des externalités et des effets d'équilibre général ne peut être également absente de l'analyse, comme elle ne peut l'être dans les autres domaines d'application de l'analyse coûts-bénéfices. Force est de constater que le calcul socioéconomique ne peut prendre en compte tous les effets d'équilibre général. En ce domaine, la recherche doit certes rester modeste et préciser les limites de l'exercice de construction d'un modèle plus général. De manière générale, plutôt que d'essayer de produire un tel modèle, soumis à des hypothèses invérifiables, il faut se limiter à étendre l'analyse à des marchés ou dimensions connexes au résultat principal recherché (par exemple, si une politique éducative vise principalement à accroître les perspectives futures d'emploi et de revenu, elle peut aussi avoir des effets sur la santé, sur le marché du logement, sur l'intensité de la délinquance, etc.). La prise en compte des effets externes doit être justifiée notamment par des résultats antérieurs montrant – sous réserve de transposabilité [d'une époque à l'autre ou d'un pays à l'autre] - que la dimension des externalités est suffisante et que les effets sur les comportements et conditions de vie des bénéficiaires de la politique ont à leur tour des effets significatifs de dissémination ou d'éviction auprès d'autres agents.

Identifier les coûts à long terme et les coûts évités

L'étape suivante, tout aussi importante, consiste à réaliser la même démarche pour les coûts à court et plus long terme. La recension des coûts est une étape cruciale dans laquelle les analystes doivent mobiliser toutes les sources et relais leur permettant d'estimer le plus précisément possible la nature et le montant de ces coûts. Cela nécessite notamment d'avoir accès aux informations qui permettent d'établir une estimation des frais en personnel et en équipements, des dépenses consacrées aux expérimentations et au suivi statistique des populations ciblées, etc. Mais il faut aussi pouvoir évaluer les coûts évités, qui sont l'un des volets les plus importants d'une politique d'investissement social, même s'ils ne sont pas spécifiques à ce type d'investissement. En effet, une telle politique est par définition susceptible d'agir à terme, par exemple, sur l'état de santé ou sur la propension à la délinquance d'une personne. Il importe donc en ce cas de pouvoir estimer le montant des dépenses de santé ou des coûts de poursuite judiciaire que la mesure permettra d'éviter. Ce n'est pas toujours chose simple, mais l'accès à des tables de valeurs correspondant à ces différents types de dépenses (pathologies, délits, etc.) devrait permettre de faire des progrès dans ce domaine. Ces tables devraient être produites avec le concours des services publics et des directions administratives concernées (Santé, Affaires sociales, Justice, etc.).

Dans un même ordre d'idées, il est nécessaire de prendre en compte les recettes fiscales et moindres transferts sociaux qui accompagnent l'augmentation potentielle des revenus des bénéficiaires, lorsque la mesure prise a ce type d'effet. Le gain salarial participe en ce sens à la réduction des coûts engendrés par la mesure.

Choisir des valeurs pour les paramètres affectant le calcul

Le calcul socioéconomique est tributaire des choix possibles pour les valeurs des paramètres imputant les coûts et les bénéfices futurs d'une mesure. Parmi ceux-ci figurent le taux d'actualisation, le taux de croissance de la productivité, le coefficient de conversion d'une amélioration des résultats scolaires en gains salariaux, ou bien encore l'horizon temporel du calcul. Beaucoup de progrès peuvent encore être faits dans ces domaines. Sur chacun de ces points, on peut d'ores et déjà formuler des propositions :

1. Il est recommandé de retenir pour le taux d'actualisation la valeur de 4,5 % préconisée dans le *Guide de l'évaluation socioéconomique des investissements publics* (2017). En général, les analyses coûts-bénéfices consacrées aux politiques d'emploi et d'éducation retiennent un taux d'actualisation annuel variant de 3 % à 6 % : ces valeurs peuvent être utilisées pour conduire des exercices de robustesse. Certaines analyses coûts-bénéfices présentent toutefois des scénarios alternatifs fondés sur des valeurs assez différentes. Ainsi, dans certaines analyses relatives aux politiques de santé, notamment en Grande-Bretagne, le taux d'actualisation préconisé est souvent inférieur à 3 %, ce qui signifie que les chercheurs accordent dans ces cas une plus grande préférence aux périodes futures.
2. Le taux de croissance des salaires individuels futurs est en général choisi en fonction des prévisions de croissance de la productivité ; les rendements salariaux de l'éducation sont dépendants du niveau de diplôme atteint ; si David Card (2001) estime que le gain salarial moyen d'une année supplémentaire d'éducation est compris entre 6 % et 10 %, Maguain (2007) estime qu'en France, ce rendement varie fortement en fonction du niveau d'éducation atteint, et qu'il est par exemple de 12 % dans le cas des études supérieures ; face à ces incertitudes, des analyses complémentaires doivent être conduites afin d'estimer la valeur actuelle du rendement de l'éducation en France ; ces analyses doivent être menées en tenant compte des niveaux et filières d'enseignement, mais aussi en intégrant les effets d'équilibre général qui peuvent se manifester ici par une modification des positions dans les « files d'attente », les nouveaux diplômés passant devant les moins diplômés dans la liste des candidatures aux emplois vacants ; en l'absence de données longitudinales de suivi sur les cohortes de bénéficiaires, les analystes peuvent néanmoins utiliser les salaires observés en coupe dans les enquêtes de type *Labour Force Surveys* (enquêtes sur l'emploi en France, *Current Population Surveys aux États-Unis*), ou bien dans des fichiers tels que les DADS en France ou le PSID aux États-Unis ;
3. Il existe encore trop peu d'études qui estiment le taux de conversion entre un accroissement des résultats à des tests scolaires standardisés et une augmentation des salaires futurs ; la recherche doit progresser sur ce point ; les quelques études consacrées à ce thème proposent de faire croître de 5 % à 15 % les salaires futurs lorsque les résultats scolaires s'améliorent d'un écart-type. Pour une amélioration inférieure, il serait convenable d'appliquer une interpolation linéaire, par exemple un gain salarial compris entre 2,5 % et 7,5 % pour un accroissement d'un demi écart-type ;

4. S'agissant des politiques d'emploi, que celles-ci soient évaluées au moyen d'expérimentations randomisées ou à l'aide d'observations quasi expérimentales, les durées d'observation des groupes cible et témoin devraient être au moins de trois ans ; dans le cas des politiques éducatives, le suivi des bénéficiaires devrait être effectué au moins pendant les quatre ou cinq années suivant la mise en œuvre de l'intervention considérée (ce qui correspond généralement à la durée d'un cycle scolaire). Si les analystes en avaient la possibilité, le suivi direct devrait être beaucoup plus long. Pour les politiques de santé, de nombreux analystes, dont le CPB aux Pays-Bas, conseillent de prendre pour horizon temporel la date de décès.

Mieux articuler évaluations *ex ante* et *ex post*

Ces conclusions nous amènent à faire deux remarques. La première est que l'évaluation socioéconomique d'une politique d'investissement social est une entreprise fastidieuse, exigeante en termes de données, de temps, de ressources, et qui conduit l'analyste à faire à chaque étape des choix souvent cruciaux et pour lesquels il n'existe pas encore de règle de décision unique et consensuelle. Des évaluations socioéconomiques *ex ante* sont utiles car elles aident à calibrer et à concevoir les politiques d'investissement social, mais elles comportent par définition une marge d'incertitude qui peut être importante. En conséquence, elles doivent être révisées au plus tôt à l'aide d'évaluations *ex post* prévues lors de la phase d'élaboration de la mesure, grâce à l'évaluation *ex ante* notamment, et conçues de sorte à recueillir les données pertinentes. C'est l'articulation de l'évaluation *ex ante* et *ex post* qui devient alors l'élément central de l'évaluation. La première ayant fourni des chiffrages initiaux que la seconde permet ou non de valider. La seconde remarque découle de la première : face à l'incertitude inhérente aux évaluations *ex ante*, et en l'absence d'un consensus sur la spécification des modèles et la valeur de leurs paramètres, il apparaît nécessaire de centraliser la production d'évaluations socioéconomiques *ex ante* selon une forme de mutualisation institutionnelle à définir (agence ou autre structure interministérielle...), en capitalisant le cas échéant sur les structures existantes. Ces travaux seront ainsi produits à partir d'une méthode unifiée, pour laquelle un consensus aura été préalablement obtenu, et qui pourra être révisée de manière transparente à la suite de nouveaux échanges, débats et contributions. Le modèle ainsi construit pourra être enrichi par des méta-analyses ou des revues systématiques de littérature permettant de dresser l'état des connaissances pour chaque politique envisagée.

Le calcul socioéconomique appliqué aux politiques d'investissement social a connu ces dernières années un regain d'intérêt, principalement parce qu'il a pu être arimé aux avancées méthodologiques accomplies dans le domaine de l'évaluation d'impact. Il est probable que les progrès qui restent à accomplir dans sa conception, sa capacité à prendre en compte une multitude d'effets, notamment à long terme, et sa précision, seront conséquents, et qu'ils constituent en cela un défi pour la recherche en sciences sociales. Le temps est peut-être venu d'accomplir la « révolution des coûts et des bénéfices » qu'appelle de ses vœux Cass Sunstein, professeur de droit à l'Université de Harvard et ancien administrateur de l'OIRA (*Office of Information and Regulatory Affairs*) du gouvernement Obama de 2009 à 2012²¹⁰.

²¹⁰ Sunstein C. (2018), *The Cost-Benefit Revolution*, MIT Press.

Résumé des conclusions

- I. Le calcul socioéconomique est étroitement lié à la pratique de l'évaluation, notamment des méthodes d'évaluation d'impact.**
- II. Il doit reposer sur des estimations robustes des effets à court et moyen terme de la politique considérée par le biais :**
 1. D'évaluations expérimentales ou quasi expérimentales,
 2. De techniques de microsimulation,
 3. De revues extensives de littérature et de méta-analyses qui peuvent fournir des ordres de grandeur pour certains paramètres du calcul.
- III. Les effets des investissements sociaux sont à rechercher dans de multiples dimensions, toutes reliées à l'activité et aux conditions de vie des individus, qu'ils soient ou non bénéficiaires de la politique évaluée :**
 1. Il s'agit ici tout à la fois d'effets potentiels sur les salaires, sur les transferts sociaux, sur les dépenses de santé, mais aussi, par exemple, sur une moindre propension à la délinquance, au recours à l'aide sociale, à une plus grande participation à la vie citoyenne, etc.
 2. Ces effets sont souvent observés dans le long terme :
 3. Le suivi de long terme des individus concernés est donc nécessaire ;
 4. Pour cela, il faut pouvoir disposer de sources d'informations statistiques longitudinales provenant de différentes sources et susceptibles d'être appariées. Les effets qui ne peuvent être quantifiés doivent néanmoins être décrits de façon qualitative.
- IV. Le calcul socioéconomique appliqué à une politique d'investissement social permet un pilotage plus fin de cette politique : l'accumulation en temps quasi-continu des estimations des effets pourrait permettre de reconfigurer cette politique ou de la déployer différemment dans un délai assez rapide.**
- V. Les investissements sociaux ont des effets hétérogènes pour des groupes différents d'individus : mesurer cette hétérogénéité des effets devrait permettre de réviser les conditions d'éligibilité à la politique considérée et mieux cibler celle-ci.**
- VI. Afin de tenir compte des externalités et des effets d'équilibre général des investissements sociaux, il est nécessaire d'étendre l'analyse à des marchés ou dimensions connexes au principal résultat recherché :**
 1. Par exemple, si l'on évalue une politique éducative, il faut pouvoir en mesurer les effets à long terme sur le marché du travail, mais aussi sur le marché du logement, la santé, la délinquance ;
 2. La prise en compte des effets externes doit être justifiée notamment par des résultats antérieurs.

- VII. Il est primordial de mobiliser toutes les sources et relais permettant d'estimer précisément la nature et le montant des coûts de la politique considérée :**
1. Cela implique d'avoir accès aux informations qui permettent d'établir une estimation des frais en personnel et en équipements, des dépenses consacrées aux expérimentations et au suivi statistique des populations ciblées, etc.
 2. Cela nécessite aussi d'évaluer les coûts évités par la mise en place de cette politique (coûts évités en termes de chômage, de délits, de mauvais état de santé, etc.)
 3. Il importe de construire des tables de valeurs correspondant à ces différents types de dépenses.
- VIII. En parallèle, il faut prendre en compte les recettes fiscales et autres transferts sociaux qui accompagnent l'augmentation potentielle des revenus des bénéficiaires**
- IX. L'analyse doit clairement évoquer les choix possibles pour les valeurs des paramètres affectant les coûts et les bénéfices futurs d'une politique : ces paramètres incluent le taux d'actualisation, le taux de croissance de la productivité, le coefficient de conversion d'une amélioration des résultats scolaires en gains salariaux, l'horizon temporel du calcul.**

Recommandations

Établir des valeurs tutélaires pour standardiser les pratiques

En vue d'appliquer avec rigueur le calcul socioéconomique aux politiques d'investissement social, il importe en premier lieu d'attribuer des valeurs monétaires aux différents effets de ces politiques, qu'il s'agisse de coûts ou de bénéfices, de dépenses induites ou le plus souvent évitées (par exemple, des pathologies, des délits, etc.). Pour toute ACB, le *Washington State Institute for Public Policy* préconise par exemple d'utiliser les valeurs tutélaires suivantes :

- La valeur d'un diplôme du secondaire (tel que le baccalauréat),
- La valeur d'un diplôme de l'enseignement supérieur,
- La valeur d'un test scolaire standardisé,
- La valeur d'une année supplémentaire d'études (en distinguant études secondaires et études supérieures),
- Le coût pour la société de la consommation d'alcool, de drogues et de tabac,
- Le coût pour la société des soins de santé physique (obésité, diabète, hypertension, etc.)
- Le coût pour la société des troubles psychiatriques et mentaux,
- Le coût pour la société d'une maltraitance subie aux âges jeunes,
- Le coût pour la société de chaque type de délit et de crime.

Dans cette optique, il serait souhaitable qu'un groupe de travail réalise, avec le concours des services publics et des directions administratives concernées (Santé, Affaires sociales, Justice, etc.), des tables attribuant des valeurs monétaires aux différents types de dépenses (pathologies, délits, etc.) engendrées par les politiques d'investissement social.

Investir en priorité le champ de l'éducation

Pour y parvenir, il serait préférable d'engager un chantier prioritaire sur le rendement social des politiques éducatives puisqu'il s'agit à notre connaissance du champ le plus mature en termes d'évaluations d'impact de qualité, dont les résultats peuvent être trouvés dans des analyses de politiques françaises et dans des revues systématiques internationales. À partir des estimations des « effets » de différentes politiques éducatives, il serait ainsi nécessaire de construire une matrice de passage entre les effets mesurés à différents âges, ou sur différentes mesures de résultats scolaires (notes à un examen ou probabilité de passage d'une classe à l'autre, par exemple). Ces matrices ne doivent pas être de simples coefficients de passages mais, étant estimées, elles devront comporter une mesure d'incertitude (écart-type). Il est par ailleurs crucial de lier ces résultats à ceux observés sur le marché du travail (taux de salaire et probabilité d'emploi notamment), sur l'état de santé (on pourra se limiter à une mesure de prévalence de la morbidité et de risque de décès prématuré, qui sera ensuite monétisée

par la valeur statistique d'une vie humaine) et, si c'est possible, sur la délinquance. C'est probablement sur ce dernier point qu'il existe moins de données et que le risque politique est le plus grand.

Pour obtenir la valeur des diplômes en termes de gain salariaux, il s'agira de construire des trajectoires de cycle de vie différenciées par niveau de diplôme, par type de diplôme, etc. Un chantier a déjà été lancé par France Stratégie en ce domaine. Les valeurs monétaires de l'état de santé devront comparer les coûts des parcours de soin et la prévalence de la morbidité et les relier au niveau d'éducation. Celles de la délinquance sont les plus compliquées à calculer en l'état, car elles supposent de mesurer les coûts directs pour le système judiciaire, les coûts des dommages infligés par types de délit, les coûts pour les victimes et leurs ayant droit, la probabilité de récidive, etc.

Pour évaluer *ex post* des politiques éducatives menées en France (comme l'emblématique réduction de la taille des classes en CP-CE1 dans le réseau d'éducation prioritaire), il convient de poursuivre le suivi longitudinal de l'échantillon d'élèves constitué par la DEPP tout au long de sa scolarité et au-delà. Dans un premier temps, pendant que ces enfants progresseront dans leur scolarité, des évaluations *ex-ante* utilisant des projections pourront être réalisées. À mesure que de nouvelles données sont produites, ces calculs *ex ante* pourront être actualisés et comparés avec les estimations précédentes. Cette politique pourrait être la première évaluation d'un investissement social de grande ampleur en France. Il faudrait toutefois être attentif à un effet potentiel qui, s'il se manifeste, serait délicat à mesurer, à savoir les externalités de cette politique sur les générations non affectées par la réforme. En effet, puisque l'augmentation du nombre d'enseignants n'a pas été parfaitement proportionnelle à l'ouverture de classes dédoublées, il est possible qu'une partie de ces enseignants soit redéployée et que, de ce fait, les générations plus âgées se soient retrouvées dans des classes plus chargées. Cet effet ne doit pas être négligé, au risque de surestimer l'effet réel de la politique.

Développer et diffuser une culture et une maîtrise de l'analyse coûts bénéfiques

Il importe aussi de favoriser le développement et la diffusion de la connaissance des méthodes d'analyse coûts-bénéfiques, à la fois dans la sphère administrative et extra-administrative, en partenariat avec les institutions universitaires et extra-universitaires qui œuvrent en ce domaine. L'objectif est ici d'initier une réflexion sur les bonnes pratiques d'évaluation, notamment les outils de microsimulation, dans un contexte de retard et de faiblesse de la force de travail sur ce thème dans notre pays. À l'instar de l'évaluation du CICE pilotée par France Stratégie, les administrations concernées devraient lancer des appels d'offres publics pour la réalisation d'ACB (*ex-ante* ou *ex-post*) et mettre en concurrence les prestataires potentiels (publics ou privés) en se souciant d'égaliser leurs chances d'accès aux données administratives.

Favoriser le développement de revues systématiques de littérature et de méta-analyses

Il est nécessaire de développer la connaissance et la maîtrise des outils de méta-analyse dans les centres de recherche et institutions dédiées. Cette recommandation fait suite au

constat de l'absence d'enseignements consacrés à ce sujet dans les universités françaises. Les revues systématiques de littérature, telles que les produisent la [Campbell Collaboration](#) ou le [CPB Netherlands Bureau for Economic Policy Analysis](#), sont un préalable indispensable à la conception d'une politique d'investissement social. Mais de telles revues ne sont pas suffisantes : elles doivent ensuite déboucher sur des méta-analyses qui permettent de calibrer les paramètres des simulations nécessaires à un exercice de type ACB.²¹¹

Favoriser le développement d'évaluations d'impact d'investissements sociaux

Le préalable à l'ACB est l'estimation robuste des effets de la politique considérée. Idéalement, celle-ci est obtenue grâce à des expérimentations randomisées avant même sa mise en place, tout au moins lorsque le temps et le budget le permettent. Favoriser le développement d'évaluations d'impact suppose donc d'accroître les financements consacrés à l'évaluation et à la mise en place d'un système d'information adéquat. Il faut comprendre que les données et résultats de ce type d'analyses sont des biens publics et que d'autres travaux peuvent être menés à partir de ces sources dès lors que celles-ci sont accessibles à toutes les équipes de recherche. Il est également crucial de pouvoir disposer de fichiers de données individuelles qui permettent d'identifier les groupes de bénéficiaires et les groupes témoins, et qui autorisent un suivi longitudinal des personnes sur une période suffisamment longue. Pour pouvoir mesurer des effets de natures différentes (par exemple, le niveau d'éducation, l'orientation scolaire et professionnelle, l'emploi, le recours aux prestations sociales, l'accès aux soins, l'état de santé, etc.), il est nécessaire d'organiser les conditions techniques (tiers de confiance, cryptage, anonymisation, etc.) qui facilitent l'appariement de données administratives (telles que celles collectées par Pôle Emploi, la Dares, la DEPP, la CNAF, la DREES, la DGFIP, l'ACOSS, etc.) afin de pouvoir suivre bénéficiaires et non bénéficiaires sur longue période, et ce en conformité avec le cadre juridique de la RGPD. Les équipes de recherche universitaires devraient être incitées à développer des enquêtes et à créer des bases de données originales pour lesquelles il n'existe pas d'équivalents dans la sphère de la statistique publique. De façon générale, elles devraient prendre plus fréquemment l'initiative de la constitution de sources statistiques de référence.

Enfin, à plus long terme, il nous apparaît important de mener une réflexion sur les institutions nécessaires au développement du calcul socioéconomique pour l'investissement social. Pour cela, nous recommandons de poursuivre l'analyse de l'applicabilité du calcul socioéconomique aux politiques d'investissement social, telle qu'elle a été engagée par le groupe de travail formé à l'initiative de France Stratégie. À cette fin, il conviendrait de mieux identifier les institutions et les chercheurs œuvrant dans ce domaine à l'étranger. L'objectif est ici d'initier le processus institutionnel devant déboucher sur une mutualisation de l'évaluation du rendement des politiques d'investissement social.

Initier une réflexion sur les meilleures modalités de mutualisation des fonctions d'évaluation socioéconomique des politiques sociales (dans un contexte de retard

²¹¹ Une méta-analyse est une étude qualitative et quantitative d'un ensemble de publications scientifiques ayant produit des estimations du même paramètre.

français et de faiblesse de la force de travail sur ce thème dans notre pays). Les équipes administratives correspondantes devraient collaborer avec les chercheurs ou institutions qui œuvrent en ce domaine, notamment celles utilisant des outils de microsimulation. Elles devraient avoir la charge d'impulser et de faciliter la mise en œuvre des recommandations précédentes, mais aussi de lancer des appels d'offres publics pour la réalisation d'ACB (*ex ante* ou *ex post*) des politiques sociales et éducatives ; à ces occasions, elles devront mettre en concurrence les prestataires potentiels (publics ou privés) en se souciant d'égaliser leurs chances d'accès aux données administratives, de préférence longitudinales, relatives aux bénéficiaires et aux informations sur les coûts, et en leur imposant des critères de qualité (et de contrôle *ex post* de celle-ci).

Bibliographie

Adler, M.D. (2011), *Well-Being and Fair Distribution: Beyond Cost-Benefit Analysis*, Oxford University Press.

Albouy, V. et L. Lequien (2009), "Does compulsory education lower mortality?", *Journal of Health Economics*, Vol. 28(1), p. 155-168

Aldy, J. E. et W. K. Viscusi (2015), "Labor market estimates of the senior discount for the value of statistical life", *Journal of Environmental Economics and Management*, Vol. 53(2007), p. 377–392.

Anderson P., Chisholm D. et D. C. Fuhr (2009), "Effectiveness and cost-effectiveness of policies and programmes to reduce the harm caused by alcohol", *Lancet*, Vol. 373(9682), p. 2234-46.

Andreoni, J. et C. Sprenger (2012), "Estimating time preferences from convex budgets", *American Economic Review*, Vol. 102(7), p. 3333-3356.

Angrist, J. et Pischke, J.-S. (2010), "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics", *Journal of Economic Perspectives*, Vol. 24(2), p. 3-30.

Atkinson, A. (1970), "On the Measurement of Inequality", *Journal of Economic Theory*, Vol. 2, p. 244-263.

Attanasio, O., Meghir, C. et A. Santiago (2012), "Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate Progresa", *The Review of Economic Studies*, vol. 79(1), p. 37–66.

Avenel, C., Boisson-Cohen, M., Dauphin, S., Duvoux, N., Fourel, C., Jullien, M. et Palier B. (2017), *L'investissement social : quelle stratégie pour la France ? La documentation Française*, Paris.

Avery, C. et S. Turner (2012), "Student Loans: Do College Students Borrow Too Much--Or Not Enough?", *Journal of Economic Perspectives*, Vol. 26(1), p. 165-192.

Bandiera, O., I. Barankay et I. Rasul (2009), "Social Connections and Incentives in the Workplace: Evidence from Personnel Data", *Econometrica*, Vol. 77(4), p. 1047-94.

Bates, M. A. et Glennerster, R. (2017), "The Generalizability Puzzle", *Stanford social innovation review*, Summer , p. 50-54.

Bayer, P., F. Ferreira et R McMillan (2007), "A unified framework for measuring preferences for schools and neighborhoods", *Journal of Political Economy*, Vol. 115(4), p. 588–638.

- Becker, G., M. DeGroot et J. Marschak (1964), "Measuring utility by a single-response sequential method", *Behavioral Science*, Vol. 9(3), p. 226–232.
- Belfield, C. R., M. Nores, W. S. Barnett et L. Schweinhart (2006), "The High/Scope Perry Preschool program: Cost-benefit analysis using data from the age-40 follow-up", *Journal of Human Resources*, Vol. 41 (1), p. 162-190.
- Bell, B., R. Costa et S. Machin (2018), "Why does education reduce crime?", *IZA Discussion Paper* No. 11805, Bonn.
- Benos, N. et S. Zotou (2014), "Education and Economic Growth: A Meta-Regression Analysis", *World Development*, Vol. 64, p. 669-689.
- Berry, J., G. Fischer et R. Guiteras (2012), *Eliciting and utilizing willingness to pay: Evidence from water filters in Ghana*, International Growth Center.
- Blanchet, D., Hagneré, C., Legendre, F. et Thibault, F. (2016), « Évaluation des politiques publiques, ex post et ex ante : L'apport de la microsimulation », *Revue économique*, Vol. 67(4), p. 685-696.
- Boiteux M. (1994), *Transports : pour un meilleur choix des investissements. Rapport pour le Commissariat Général du Plan*, La Documentation Française, Paris, 132 p.
- Boiteux M. (2001), *Transports : choix des investissements et coût des nuisances, Rapport pour le Commissariat Général du Plan*, La Documentation Française, Paris, 328 p.
- Bono, P.-H., S. Debu, R. Desplatz, M. Hayet, C. Lacouette-Fougère, et A. Trannoy (2018), « Vingt ans d'évaluations d'impact en France et à l'étranger. Analyse quantitative de la production scientifique », *Document de travail*, n°2018-06, décembre, France Stratégie.
- Bouguen, A., J. Grenet, et M. Gurgand (2017), « La taille des classes influence-t-elle la réussite scolaire ? », Note n°28 de l'I.P.P., Paris.
- Bozio A. et P.Y Geoffard (2017), *L'accès des chercheurs aux données administratives - Etat des lieux et propositions d'actions. Rapport pour le Secrétariat d'Etat à l'industrie, au numérique et à l'innovation*, Conseil national de l'information statistique, La Documentation Française, Paris, 68 pages.
- Brady, H. B., M. Hout et J. Stiles (2005), *Return on Investment: Educational Choices and Demographic Change in California's Future*, Survey Research Center, University of California Berkeley
- Brodeur, A., Cook, N. et Heyes, A. (2018), "Methods matter: P hacking and causal inference in economics", *IZA Discussion paper* No. 11796, Bonn.
- Burgoon, B. (2015), "Practical pluralism in the empirical study of social investment: Examples from active labour-market policy", dans *The Uses of Social Investment*, édité par A. Hemerijck, Oxford: Oxford University Press, p. 161-173.
- Cahuc, P., S. Carcillo et T. Le Barbanchon (2017), "The Effectiveness of Hiring Credits", *CEPR Discussion Paper* no. 12537, à paraître dans *The Review of Economic Studies*.
- Campbell Collaboration (2018), *Annual report 2017*, Oslo.
- Card, D. (2001), "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems", *Econometrica*, Vol. 69(5), p. 1127-1160.

- Card, D., Kluve, J. et Weber, A. (2018), "What Works? A Meta-Analysis of Recent Active Labor Market Program Evaluations", *Journal of the European Economic Association*, Vol 16(3), p. 894-931.
- Carneiro, P. et R. Ginja (2014) "Long-Term Impacts of Compensatory Preschool on Health and Behavior: Evidence from Head Start", *American Economic Journal: Economic Policy*, Vol. 6, p. 135-73.
- Cellini, S. R. et Kee, J. E. (2015), "Cost effectiveness and cost benefit analysis", dans *Handbook of Practical Program Evaluation*, édité par K. Newcomer, H. Hatry et J. Wholey, 4ème édition, Jossey-Bass, p. 493-530.
- Chamot, E., Bovier, P., Perneger, T., Chastonay, P., Herrmann, F., Garnerin, P. et Chopard, P. (2001), « Analyse économique appliquée à la santé : un tour d'horizon », *Bulletin des médecins suisses*, Vol. 82, n°39.
- Chassang, S., G. Padro, I. Miquel et E. Snowberg (2012), "Selective trials: A principal-agent approach to randomized controlled experiments", *American Economic Review*, vol. 102(4), p. 1279-1309.
- Chetty, R., A. Guren, D. Manoli et A. Weber (2012), "Does indivisible labor explain the difference between micro and macro elasticities? A meta-analysis of extensive margin elasticities", *NBER Macroeconomics Annual 2012*, Vol. 27(1), p. 1-56.
- Chetty, R., J. Friedman, N. Hilger, E. Saez, D. Whitmore Schanzenbach et D. Yagan (2011), "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star", *The Quarterly Journal of Economics*, Vol. 126(4), p. 1593-1660.
- Cheung S. L. (2016), "Recent developments in the experimental elicitation of time preference", *Journal of Behavioral and Experimental Finance*, Vol. 11, p. 1-8.
- Chevassus-au-Louis B. (2009), *Approche économique de la biodiversité et des services liés aux écosystèmes*, Rapport pour le Conseil d'analyse stratégique, La Documentation Française, 376 p.
- Clark, D. et H. Royer (2013), "The Effect of education on adult mortality and health: Evidence from Britain", *American Economic Review*, Vol. 103(6), p. 2087-2120.
- Cobiac, L., T. Vos et J. J. Barendregt (2009), "Cost-effectiveness of interventions to promote physical activity: a modelling study", *PLoS Medicine*, Vol. 6(7), e1000110.
- Cobiac, L., T. Vos et J. L. Veerman (2010), "Cost-effectiveness of interventions to promote fruit and vegetable consumption", *PloS one*, Vol. 5(11), e14148.
- Cobiac, L., T. Vos, C. Doran et A. Wallace (2009), "Cost-effectiveness of interventions to prevent alcohol-related disease and injury in Australia", *Addiction*, Vol. 104(10), p. 1646-1655.
- Cohen, M.A. (2005), *The Costs of Crime and Justice*, Routledge, New York.
- Commission européenne (2015), « L'investissement social : définition », <http://ec.europa.eu/social/main.jsp?catId=1044&langId=fr>
- Conti, G., J. Heckman et R. Pinto (2016), "The effects of two influential early childhood interventions on health and healthy behavior", *The Economic Journal*, Vol. 126(596), p. 28-65.

Connolly, P., C. Keenan et K. Urbanska (2018), "The trials of evidence-based practice in education: a systematic review of randomised controlled trials in education research 1980–2016", Vol. 60(3), *Educational Research*, p. 276-291.

Crépon, B., E. Duflo, M. Gurgand, R. Rathelot et P. Zamora (2013), "Do Labor Market Policies have Displacement Effects? Evidence from a Clustered Randomized Experiment", *The Quarterly Journal of Economics*, Vol. 128(2), p. 531–580.

Crépon, B., Gurgand, M., Kamionka, T. et Lequien, L. (2012), "Is counseling welfare recipients cost effective? Lessons from a randomized experiment", Working Paper, J-Pal, Paris School of Economics.

Dalkey, N. C. et O. Helmer (1963), "An experimental application of the Delphi Method to the use of experts", *Management Science*, Vol. 9(3), p. 458-468.

Dalziel, K. M., D. Halliday et L. Segal (2015), "Assessment of the Cost–Benefit Literature on Early Childhood Education for Vulnerable Children: What the Findings Mean for Policy", *Sage Open*, Vol. 5(1).

Damon, J (2015), « L'investissement social : contenu et portée d'une notion en vogue », *Revue de droit sanitaire et social*, n° 4, p. 722-733.

Deming, D. (2009) "Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start", *American Economic Journal: Applied Economics*, vol. 1, p. 111-34.

Desplatz, R. et Ferracci, M. (2016), Comment évaluer l'impact des politiques publiques ? Un guide à l'usage des décideurs et praticiens, Document de travail, France stratégie.

Dhaliwal, I., E. Duflo, R. Glennerster et C. Tulloch (2013), "Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries: A General Framework with Applications for Education", dans *Education Policy in Developing Countries*, University of Chicago Press, p. 285-338.

Dhami S. (2016), *The Foundations of Behavioral Economic Analysis*, Oxford Economic Press.

Diamond, P. et J. Hausman (1994), "Contingent Valuation: Is Some Number Better than No Number?", *Journal of Economic Perspectives*, Vol. 8(4), p. 45-64.

Drèze, J. et N. Stern (1990), "Policy reform, shadow prices, and market prices", *Journal of Public Economics*, Vol. 42(1), p. 1-45.

Dynarski, S. (2004), "The new merit aid", dans *College Choices: The Economics of Where to Go, When to Go, and How To Pay for It*, édité par C. Hoxby, University of Chicago Press.

Fack, G. et J. Grenet (2015), "Improving College Access and Success for Low-Income Students: Evidence from a Large Need-Based Grant Program", *American Economic Journal: Applied Economics*, Vol.7(2), p. 1-34.

Filatrou, O., D. Fougère et M. Tô (2013), "Will Sooner Be Better? The Impact of Early Preschool Enrollment on Cognitive and Noncognitive Achievement of Children", Discussion Paper No 9480, C.E.P.R., Londres.

Filges T., Jonassen A. B., Jørgensen A. M. K. (2018), "Reducing unemployment benefit duration to increase job finding rates: a systematic review", *Campbell Systematic Reviews* 2018:2

Filges T., G. Smedslund, A. S. D. Knudsen et A. M. K. Jørgensen . (2015), "Active Labour Market Programme Participation for Unemployment Insurance Recipients: A Systematic Review", *Campbell Systematic Reviews*, 2015:2.

Filges, T., Sonne-Schmidt, C. S. et Nielsen, B. C. V. (2018), "Small class sizes for improving student achievement in primary and secondary schools", *Campbell Systematic Reviews*, 2018:10

Fleurbaey, M. (2008), *Fairness, Responsibility, and Welfare*. Oxford University Press.

Fleurbaey, M. (2010) Defense of the ex-post evaluation of risk, Note.

Fleurbaey, M. et M. D. Adler. (2016), *The Oxford Handbook of Well-Being and Public Policy*, Oxford University Press

Fleurbaey, M., S. Luchini, E. Schokkaert et C. Van de Voorde (2012), « Évaluation des politiques de santé : pour une prise en compte équitable des intérêts des populations », *Economie et Statistique*, n° 455-456, p. 11-36.

Fleurbaey, M. et V. Peragine (2013), "Ex Ante Versus Ex Post Equality of Opportunity", *Economica* Vol. 80, p.118–130

France Stratégie et DG Trésor (2017), *Guide de l'évaluation socioéconomique des investissements publics*.

France stratégie (2019), *L'évaluation socioéconomique des projets immobiliers dans l'enseignement supérieur et la recherche*, sous la direction d'Emile Quinet, rapporteur, J. Ni, Paris.

France stratégie (2017), The discount rate in the evaluation of public policy, Proceedings of the conference organized by Commissariat général à l'investissement Conseil général de l'environnement et du développement durable France Stratégie, Paris.

Fredriksson, P., B. Öckert et H. Oosterbeek (2013), "Long-Term Effects of Class Size", *The Quarterly Journal of Economics*, Vol. 128(1), p. 249–285.

Fryer, R.G. Jr. (2016). "The Production of Human Capital in Developed Countries: Evidence from 196 Randomized Field Experiments », NBER Working Papers 22130, National Bureau of Economic Research, Inc.

Fullerton, D. et G. Heutel (2011), "Analytical general equilibrium effects of energy policy on output and factor prices", *The B.E. Journal of Economic Analysis et Policy*, Vol. 10(2), p. 1-26.

Galinié, A. et A. Heim, (2016), « Inégalités scolaires : quels rôles jouent les cours privés ? » Contribution au rapport du Cnesco Les inégalités scolaires d'origines sociales et ethnoculturelle, Paris : Cnesco.

Garces, E., D. Thomas et J. Currie (2002), "Longer-Term Effects of Head Start", *American Economic Review*, Vol. 92, p. 999-1012.

García, J. L., J. J. Heckman, D. E. Leaf et M. J. Prados (2017), "Quantifying the Life-cycle Benefits of a Prototypical Early Childhood Program", *NBER Working Paper* No. 23479.

GBD Risk Factors Collaborators (2016), "Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015", *Lancet*, 388, p. 1659-724

Gneezy, U. et A. Rustichini (2000), "A fine is a price", *Journal of Legal Studies*, Vol. 29(1), p. 1-18.

Gollier C. (2011), Le calcul du risque dans les investissements publics. Rapport pour le Conseil d'Analyse Stratégique, La Documentation Française, 240 p.

Goulder, L. et R. Williams (2003), "The substantial bias from ignoring general equilibrium effects in estimating excess burden, and a practical solution", *Journal of Political Economy*, Vol. 111(4), p. 898-927.

Gurevitch, J., Koricheva, J., Nakagawa, S. et Stewart, G. (2018), "Meta-analysis and the science of research synthesis", *Nature*, Vol. 555, Mars, p. 175-182.

Haute Autorité de santé (2014), « Valeurs de références pour l'évaluation économique en santé. Revue de la littérature », Document de travail, HAS.

Heckman, J. J., L. Lochner et C. Taber (1998), "General-equilibrium treatment effects: A study of tuition policy", *American Economic Review*, Vol. 88(2), p.381-386.

Heckman, J. J., J. Humphries et G. Veramendi (2017), "The Non-Market Benefits of Education and Ability", *NBER Working Paper* No 23896.

Heckman, J. J., N. Hohmann, J. Smith et M. Khoo (2000), "Substitution and Dropout Bias in Social Experiments: A Study of an Influential Social Experiment", *The Quarterly Journal of Economics*, Vol. 115(2), p. 651-694.

Heckman, J. J., S. H. Moon, R. Pinto, P. Savelyev et A. Yavitz (2010), "The rate of return to the High Scope Perry Preschool Program," *Journal of Public Economics*, Vol. 94(1-2), p. 114-128.

Heim, A (2017a), « Comment estimer le rendement de l'investissement social ? », *Document de travail* 2017-02, France stratégie.

Heim, A (2017b), « Peut-on estimer le rendement de l'investissement social ? », *Note d'analyse* n° 52, France stratégie.

Jenson, J. (2010), "Diffusing ideas for after neoliberalism: The social investment perspective in Europe and Latin America", *Global Social Policy*, Vol. 10(1), p. 59-84.

Jepsen, C. et S. Rivkin (2002), "Class Size Reduction, Teacher Quality, and Academic Achievement in California Elementary Public Schools", San Francisco: Public Policy Institute of California.

Jepsen, C. et S. Rivkin (2009), "Class Size Reduction and Student Achievement: The Potential Tradeoff between Teacher Quality and Class Size", *The Journal of Human Resources*, Vol. 44(1), p. 223-250.

- Kaas, K. P. et H. Ruprecht (2006), "Are Vickrey auction and the BDM mechanism really incentive compatible? Empirical results and optimal bidding strategies in cases of uncertain willingness to pay", *Smalenbach Business Review*, Vol. 58(1), p. 37-55.
- Kah, E. (2003), « La méthode d'évaluation contingente appliquée aux déchets urbains », *L'Espace géographique*, Vol. 32(1), p. 47-59.
- Kaplan, R.M. et Irvin V. L. (2015), "Likelihood of Null Effects of Large NHLBI Clinical Trials Has Increased over Time", *PLoS ONE*, Vol. 10(8): e0132382
- Kirkeboen, L. J., E. Leuven, et M. Mogstad (2016), "Field of study, earnings and self-selection", *Quarterly Journal of Economics*, Vol. 131(3), p. 1057–1111.
- Klaiber, H. A. et V.K. Smith (2010), "General equilibrium benefit analyses for social programs", in *Towards Principles and Standards in the Use of Benefit-Cost Analysis: A Summary of Work, and a Starting Place*, édité par R. Zerbe et alii, Seattle, WA: Benefit-Cost Analysis Center, University of Washington.
- Klaiber, H.A. et V.K. Smith (2012), "Developing General Equilibrium Benefit Analyses for Social Programs: An Introduction and Example", *Journal of Benefit-Cost Analysis*, Vol. 3(2), p. 1-50.
- Kline, P. et C. Walters (2016), "Evaluating Public Programs with Close Substitutes: The Case of Head Start", *The Quarterly Journal of Economics*, Vol. 131(4), p. 1795–1848.
- Kluve J., Puerto S., Robalino D., Romero J. M., Rother F., Stöterau J., Weidenkaff F., Witte M. (2017), "Interventions to improve the labour market outcomes of youth: a systematic review of training, entrepreneurship promotion, employment services, and subsidized employment interventions", *Campbell Systematic Reviews* 2017:12.
- Knai C, Pomerleau J, Lock K, McKee M. Getting children to eat more fruit and vegetables: a systematic review. *Prev Med.* 2006 ; Vol. 42(2) ; p. 85-95.
- Kok, L., C. Koopmans, C. Berden, et R. Dosker (2011), « De waarde van kinderopvang », Amsterdam: SEO (SEO Economisch Onderzoek, No. 2011-29.
- Kriström, B. (2006), "Framework for assessing the distribution of financial effects of environmental policies", dans Y. Serret et N. Johnstone (éditeurs), *The Distributional Effects of Environmental Policy*, Cheltenham, Edward Elgar et OCDE, Paris.
- Krueger, A. (1999), "Experimental Estimates of Education Production Functions", *The Quarterly Journal of Economics*, Vol. 114(2), p. 497-532.
- Krueger, A. et D. Whitmore (2001), "The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR", *The Economic Journal*, vol. 111(468), p. 1–28.
- Kvist, J. (2015), "A framework for social investment strategies: Integrating generational, life course and gender perspectives in the EU social investment strategy", *Comparative European Politics*, Vol. 13(1), p. 131–149.
- Lacetera, N. et M. Macis (2013), "Time for blood: The effect of paid leave legislation on altruistic behavior", *Journal of Law, Economics, and Organization*, Vol. 29(6), p. 1384–1420.

Lane, T. (2016), "Discrimination in the laboratory: A meta-analysis of economics experiments", *European Economic Review*, Vol. 90, p. 375-402.

Le Gall-Ely, M. (2009), « Définition, mesure et déterminants du consentement à payer du consommateur : synthèse critique et voies de recherche », *Recherche et Applications en Marketing (French Edition)*, SAGE Publications, Vol. 24(2), p.91-113.

Lebègue D. (2005), *Révision du taux d'actualisation des investissements publics*, Rapport pour le Commissariat Général du Plan, La Documentation Française, Paris, 112 p.

Levin, H. et P. McEwan (2001). *Cost-Effectiveness Analysis*. Thousand Oaks, CA: Sage Publications Inc.

Li, R., Qu, S., Zhang, P., Chattopadhyay, S., Gregg, E. W., Albright, A., Hopkins, D. et N. P. Pronk (2015), "Economic Evaluation of Combined Diet and Physical Activity Promotion Programs to Prevent Type 2 Diabetes Among Persons at Increased Risk: A Systematic Review for the Community Preventive Services Task Force", *Annals of Internal Medicine*, Vol. 163(6), p. 452-460.

Linstone, H. et M. Turrof (1975), *The Delphi method, techniques and applications*. Addison Wesley Publishing.

Lipsey, M. W. et D. Wilson (2001), *Practical Meta-Analysis*, Sage Publisher.

Lochner L. et E. Moretti (2004), "The effect of education on crime: Evidence from prison inmates, arrests, and self-report", *American Economic Review*, Vol. 94(1), p. 155-189.

Lochner, L. (2004), "Education, work and crime: A human capital approach", *International Economic Review*, Vol. 45(3), p. 811-843.

Loomis, J. B., P. Bell, H. Zita-Cooney et C. Asmus (2009), "A comparison of actual and hypothetical willingness to pay of parents and non-parents for protecting infant health: The case of nitrates in drinking water", *Journal of Agricultural and Applied Economics*, Vol. 41(3), p. 697-712.

Loomis, J. B. (2011), "Incorporating distribution into BCA using nonmarket valuation", *Journal of Costs-Benefits Analysis*, Vol. 2(1), p. 1-24.

Low, H. et C. Meghir (2017), "The Use of Structural Models in Econometrics", *Journal of Economic Perspectives*, vol. 31(2), p. 33-58.

Machin, S. et S. McNally (2008), "The literacy hour", *Journal of Public Economics*, vol. 92(5-6), p. 1441-1462.

Maguain, D. (2007), « Les rendements de l'éducation en comparaison internationale », *Economie et prévision*, n° 180-181, p. 87-106.

Manski, Charles F. (2015). Narrow or broad cost-benefit analysis? *Criminology & Public Policy* 14(4): 647-51.

Mas, A. et E. Moretti (2009), "Peers at work", *American Economic Review*, Vol. 99(1), p. 112-145.

Milligan, K., E. Moretti et P. Oreopoulos (2004), "Does education improve citizenship? Evidence from the United States and the United Kingdom", *Journal of Public Economics*, Vol. 88(9-10), p. 1667-1695.

- Moretti, E. (2004), "Estimating the social return to higher education", *Journal of Econometrics*, Vol. 121(1-2), p. 175-212.
- Murdock, S., S. White, N. Hoque, B. Pecotte, X. You et J. Balkan (2003), *The New Texas Challenge: Population Change and the Future of Texas*. College Station: Texas A&M University Press.
- Murnane, R., J. Willett et F. Levy (1995), "The Growing Importance of Cognitive Skills in Wage Determination", *The Review of Economics and Statistics*, Vol. 77(2), p. 251-66
- Murray, C. et A. Lopez (1997), "Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study", *Lancet*, 349, p. 1436-1442.
- Neal, D. et W. Johnson (1996), "The Role of Premarket Factors in Black-White Wage Differences", *Journal of Political Economy*, Vol 104(5), p. 869-895.
- Olken, B. (2015) Promises and perils of Pre-analysis plans, *Journal of Economic Perspectives*, V. 29, N°3, p 61–80
- Palier, B. (2014), *La stratégie d'investissement social*. Les études du Conseil économique, social et environnemental, La Documentation Française, Paris, 68 pages.
- Persson, M. et M. Svensson (2013), "The willingness to pay to reduce school bullying", *Economics of Education Review*, Vol. 35, p. 1-11.
- Pol, T. V., F. Bos et G. Romijn (2016), "Distributionally weighted cost-benefit analysis: From theory to practice", CBP discussion paper 364, Den Haag.
- Puma, M. et alii, (2010) "Head Start Impact Study. Final Report", Washington DC: U.S. Department of Health and Services, Administration for Children and Families.
- Quinet A. (2008), *La valeur tutélaire du carbone*, Rapport pour le Conseil d'Analyse Stratégique, La Documentation Française, 424 p.
- Quinet E. (2013), *L'évaluation socioéconomique des investissements publics, Tome 1*, Rapport pour le Commissariat général à la stratégie et à la prospective, La Documentation Française.
- Quinet E. (2013), *L'évaluation socioéconomique des investissements publics, Tome 2*, Rapport pour le Commissariat général à la stratégie et à la prospective, La Documentation Française
- Ramos, D., P. Arezes et P. Afonso (2016), "Application of the Delphi Method for the inclusion of externalities in occupational safety and health analysis", *DYNA*, Vol. 83(196), p. 14-20.
- Reynolds, A. J., J. A. Templ, B. A. White, S. R. Ou et D. L. Robertson (2011), "Age-26 Cost-Benefit Analysis of the Child-Parent Center Early Education Program", *Child Development*, Vol. 82(1), p. 379–404.
- Rolnick, A. et R. Grunewald (2003), "Early childhood development: Economic development with a high public return", Technical Report, Federal Reserve Bank of Minneapolis, Minneapolis, MN.
- Romijn, G. et G. Renes (2013), *General Guidance for Cost-Benefit Analysis*, CPB/PBL, Den Haag.

- Salomon, J. A., Haagsma, J. A., Davis, A., de Noordhout, C. M., Polinder, S., et Havelaar, A. H. (2015), "Disability weights for the global burden of disease 2013 study", *Lancet Glob Health*, Vol.3(11), p.712-723.
- Samson A. L., E. Schokkaert, C. Thébaut, B. Dormont, M. Fleurbaey, S. Luchini et C. Van de Voorde (2018), "Fairness in cost-benefit analysis: a methodology for health technology assessment", *Health Economics*, Vol. 27, p. 102–114.
- Samuelson, P. et W. Nordhaus (2000), *Economie*. 16^e édition, Paris : Economica.
- Shih, S., R. Carter, C. Sinclair, C. Mihalopoulos et T. Vos (2009), "Economic evaluation of skin cancer prevention in Australia". *Preventive Medicine*, Vol. 49(5), p. 449-53.
- Siedler, T. (2010), "Schooling and citizenship in a young democracy: Evidence from postwar Germany", *Scandinavian Journal of Economics*, Vol. 112(2), p. 315-338.
- Stiglitz, J. (2000), "Shadow prices and market prices", dans J. Stiglitz (2000), *Economics of the public sector*, 3rd Edition, Columbia.
- Studer, N. (2011), *Méthodes d'évaluation des politiques publiques : actes du séminaire*, DREES, Paris.
- Sunstein, C. (2018), *The Cost-Benefit Revolution*, MIT Press.
- Tirole, J. (2009), « Motivation intrinsèque, incitations et normes sociales », *Revue économique*, Vol. 60(3), p. 577-589.
- Treasury of New Zealand Government (2015), *Guide to social costs-benefits analysis*.
- Van der Pol, T., F. Bos et G. Romijn (2017), "Distributionally weighted cost-benefit analysis: From theory to practice", *CPB Discussion Paper* n° 364, Centraal Planbureau, Den Haag.
- Vos, T., R. Carter, C. Mihalopoulos, L. Veerman, A. Magnus, L. Cobiac et alii (2010), *Assessing Cost-Effectiveness in Prevention*. Brisbane, Melbourne: University of Queensland and Deakin University, 101 pages.
- Washington State Institute for Policy Evaluation (2018), "Class size: reducing average class size by one student in kindergarten Pre-K", <http://www.wsipp.wa.gov/BenefitCost/Program/204>
- Weimer, D. et A. Vining (2009), *Investing in the Disadvantaged: Assessing the Benefits and Costs of Social Policies*, Georgetown University Press, 304 pages.
- Weimer, D. et A. Vining (2010), *Policy Analysis: Concepts and Practice*. Routledge, Oxford.
- Wilson, S. J., E. Tanner-Smith, M. Lipsey, K. Steinka-Fry et J. Morrison (2011), "Dropout prevention and intervention programs: Effects on school completion and dropout among school- aged children and youth", *Campbell Systematic Reviews* 2011:8.
- Zief S. G., S. Lauver et R. A. Maynard (2006), "Impacts of after-school programs on student outcomes", *Campbell Systematic Reviews* 2006:3.
- Zoutman, F. T., B. Jacobs et E. L. W. Jongen (2016), "Redistributive politics and the tyranny of the middle class", *CESifo Working Paper Series* 5881, CESifo Group Munich.

Annexes

Annexes A

Mise en pratique d'analyses coûts-bénéfices d'investissements sociaux

Les chapitres précédents ont permis de détailler les éléments et étapes nécessaires à la mise en œuvre d'analyses coûts-bénéfices d'investissements sociaux. Afin d'illustrer et clarifier nos propos, nous avons souvent fait référence à des publications académiques. Mais à ce stade, il est encore difficile d'avoir une vue précise des méthodes proposées et surtout de la façon de les mettre en œuvre concrètement dans des contextes particuliers. Dans cette annexe, nous présentons de façon relativement détaillée quatre analyses coûts-bénéfices d'investissements sociaux. Ces exemples ont pour vertu de faire apparaître plus nettement les éléments de méthode principaux et les conséquences des choix de certaines hypothèses. Les encadrés « que retenir » synthétisent ces éléments pour chacun des exemples présentés. Les articles et études ici présentés sont représentatifs des choix et des méthodes que devrait faire l'analyste souhaitant réaliser l'analyse coûts-bénéfices d'un investissement social.

Les présentations suivent un ordre allant de l'étude la plus simple, qui est une application directe du principe de calcul d'une valeur actualisée nette des flux de revenus et de coûts futurs (Krueger et Whitmore, 2001), à une formulation plus générale et plus complexe du rendement d'un investissement social, intégrant les bénéfices et les coûts directs et induits à long terme d'un investissement social (Heckman *et al.*, 2010). Sont ici résumées tout à la fois des études spécifiques, concernant une expérience ou une intervention particulière, et des méta-analyses synthétisant un ensemble de recherches portant sur un type d'intervention donné.

Les deux premières sections sont consacrées à la réduction de la taille des classes. La première résume l'étude de Krueger et Whitmore (2001), qui concerne uniquement le programme *STAR* de réduction de la taille des classes d'écoles primaires. La seconde livre une méta-analyse de cette politique à partir des études réalisées aux U.S.A. sur le même thème. Elle a été réalisée par le *Washington State Institute for Public Policy* (WSIPP) qui est un organisme spécialisé dans la production de méta-analyses de politiques sociales et éducatives. Les résultats du WSIPP présentent plusieurs intérêts. Premièrement, son analyse coûts-bénéfices tient compte des bénéfices et coûts directs mais aussi induits d'une telle politique, et ce sur le long terme (à l'échelle d'une vie humaine). Deuxièmement, le WSIPP propose une solution au problème des doubles comptes qui apparaît lorsque certains résultats ont la même cause directe, à savoir le programme évalué, et lorsqu'ils sont simultanément monétisés dans le calcul des bénéfices ; en ce cas, l'estimation risque de surestimer, ou tout au moins de mésestimer, le véritable rendement social du programme. Troisièmement, le WSIPP teste la robustesse de ses résultats à l'aide d'une analyse de sensibilité par simulation. La méta-analyse réalisée par le WSIPP étend, tout en la confirmant, celle plus limitée conduite par Krueger et Whitmore (2001).

Les deux sections suivantes sont consacrées à des politiques de santé. La première est une méta-analyse d'une centaine de politiques préventives ; elle a été réalisée par des chercheurs australiens. La suivante présente les résultats d'une recherche conduite par des économistes français sur l'efficacité des traitements de l'hypertension. Cette recherche

présente plusieurs intérêts. Premièrement, elle mobilise la méthode des préférences exprimées par les individus eux-mêmes afin de valoriser les bénéfices de ces traitements. Deuxièmement, elle repose sur la simulation de trajectoires de santé. Par ailleurs, elle permet de prendre en compte différents degrés d'aversion aux inégalités et de mesurer leur impact sur les résultats de l'évaluation.

La dernière section est consacrée à l'analyse coûts-bénéfices du *Perry Preschool Program* réalisée par Heckman *et al.* (2010). Cette ACB est l'une des plus complètes dont nous disposons à ce jour. En un sens, elle constitue « l'idéal » de ce que devrait être une application du calcul socioéconomique à une politique d'investissement social. Elle est adossée à une expérience aléatoire contrôlée et à un suivi longitudinal très long, reposant sur l'appariement des observations à des sources administratives (en particulier à des sources judiciaires permettant de valoriser les coûts de la délinquance). L'étude contient des analyses de sensibilité très détaillées, tenant compte des valeurs par nature incertaines de plusieurs paramètres.

Annexe A-I

Le programme STAR de réduction de la taille des classes

Cadre général

La réduction de la taille des classes, notamment dans les écoles élémentaires classées en zone d'éducation prioritaire, est l'une des politiques éducatives qui est supposée profiter le plus aux élèves en difficulté et limiter de ce fait les inégalités de réussite scolaire. Comme l'a rappelé l'encadré dans le chapitre II.4, les preuves quant à l'efficacité de cette mesure proviennent pour l'essentiel du succès d'une expérimentation fréquemment commentée, le projet *STAR* (*Student/Teacher Achievement Ratio*) mis en œuvre au Tennessee dans les années quatre-vingt. Dans le cadre de cette expérimentation, 11 600 élèves entrant dans l'équivalent de la grande section de maternelle en France (appelée *kindergarten* aux États-Unis), donc en moyenne âgés de 5 ans, furent aléatoirement affectés, ainsi que leurs enseignants, soit à une classe de taille habituelle (comprise entre 22 et 25 élèves), soit à une classe de taille habituelle mais proposant un enseignement plus intense, caractérisé par la présence d'un autre enseignant dans la classe, ou bien à une classe de taille plus réduite (comprise entre 13 et 17 élèves). L'affectation à l'un de ces trois types de classes était pérennisée pendant quatre ans, donc jusqu'à la fin de l'équivalent de la classe de CE2 en France (appelée *grade 3* aux États-Unis).

L'analyse statistique des effets de cette expérience sur les résultats scolaires des enfants en fin d'année scolaire, année par année, du *kindergarten* jusqu'à la fin du CE2, a été réalisée par Krueger dans un article publié en 1999 dans le *Quarterly Journal of Economics*.²¹² Les résultats de son étude sont très significatifs. Les principales conclusions en sont les suivantes : 1) à l'issue de l'année de *kindergarten*, les élèves ayant été aléatoirement affectés à des classes de taille réduite ont obtenu des résultats en moyenne plus élevés à

²¹² Krueger A. (1999), "Experimental Estimates of Education Production Functions", *The Quarterly Journal of Economics*, Vol. 114(2), p. 497-532.

des tests standardisés ²¹³, 2) l'écart en faveur des élèves affectés à des classes de taille réduite s'accroît d'année en année, ²¹⁴ 3) les soutiens apportés aux élèves dans les classes de taille habituelle n'ont que peu ou pas d'effets, et 4) la réduction de la taille de classe bénéficie plus aux enfants de milieux défavorisés.

Intérêt de la démarche

Pour faire suite à cette première analyse, Krueger, en collaboration avec Whitmore, a proposé une analyse coûts-bénéfices du projet STAR.²¹⁵ Cette seconde étude est exemplaire de plusieurs points de vue.

- Premièrement, elle est **adossée aux résultats d'une expérimentation randomisée qui produit des estimations robustes de l'effet moyen à court et moyen terme de la mesure considérée.**
- Deuxièmement, elle **met en œuvre de façon particulièrement claire les principes du calcul de la variation de la valeur actualisée nette (VAN)** pour cette intervention, principes qui ont été exposés dans le chapitre précédent.
- Troisièmement, les **limites de l'exercice sont précisément énoncées, et chaque composante du calcul est soigneusement commentée.**

Les auteurs étant conscients des incertitudes pesant sur les valeurs possibles de ces composantes, ils **ne se bornent pas à un chiffrage unique** et considèrent plusieurs scénarios relativement plausibles.

En revanche, ils ne **prennent pas en compte les effets induits de cette politique**, tant du point de vue budgétaire (*via* les gains de recettes fiscales induits par les salaires plus élevés perçus par les bénéficiaires de la réduction de la taille des classes) que sociétal (en monétisant, par exemple, les effets de cette politique sur la qualité de la vie privée et familiale des bénéficiaires, leur santé, leur moindre propension à la délinquance, leur comportement citoyen, etc.). La valorisation monétaire des effets est fondée sur une autre étude (voir *infra*) permettant de « convertir » des résultats scolaires en gains salariaux.

Néanmoins, et c'est là probablement l'aspect de l'étude qui a été le plus popularisé, Krueger et Whitmore trouvent que dans la plupart des scénarios envisagés, les bénéfices salariaux directs d'une réduction significative de la taille des classes sont plus élevés que les coûts de cette politique éducative. Ainsi, les analyses coûts-bénéfices de Krueger et Whitmore sont présentées comme « conservatrices » dans le sens où elles n'intègrent que les effets directs sur l'emploi et, malgré tout, baisser la taille des classes se révèle un investissement socialement rentable dans la plupart des estimations.

²¹³ Leurs résultats sont supérieurs de quatre centiles à ceux des élèves affectés à des classes de taille habituelle, bénéficiant ou non d'un enseignement plus intensif. Rappelons qu'en statistique descriptive, un centile, ou percentile, est chacune des 99 valeurs qui divisent les données triées en 100 parts égales, de sorte que chaque partie représente un centième de l'échantillon.

²¹⁴ Krueger (1999) estime que l'accroissement annuel moyen de cet écart est égal à un centile.

²¹⁵ Krueger A. et Whitmore D. (2001), "The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR", *The Economic Journal*, vol. 111(468), p. 1-28.

Hypothèses

Pour réaliser leur calcul, Krueger et Whitmore (2001) font plusieurs hypothèses. En premier lieu, ils supposent que les élèves entrent tous à l'école à l'âge de 5 ans (au *kindergarten*), qu'ils commencent tous à travailler à l'âge de 18 ans et prennent leur retraite à 65 ans. L'hypothèse selon laquelle personne ne fait d'études supérieures neutralise l'effet positif de celles-ci sur les salaires et fournit donc une borne inférieure pour l'évaluation des bénéfices salariaux de la mesure. Dans le cas du projet STAR, les effectifs ont été réduits en moyenne de sept élèves par classe et la taille moyenne d'une « petite » classe était de quinze élèves. En conséquence, le coût annuel d'un élève affecté à une classe de taille réduite est supérieur de 47 % (soit 7 divisé par 15) à celui d'un élève scolarisé dans une classe de taille standard. Krueger et Whitmore font l'hypothèse que celui-ci est égal au coût moyen d'un élève scolarisé aux États-Unis en 1997-1998, soit 7 502 dollars.

En théorie, chaque élève participant à l'expérience STAR devait être affecté à une classe de taille constante pendant quatre ans (du *kindergarten* au *grade 3*). Mais en raison des changements d'établissements, on a observé que les élèves étaient en moyenne restés seulement 2,3 années dans une classe de même taille. En conséquence, Krueger et Whitmore ont fait l'hypothèse que les coûts de la mesure étaient entièrement encourus durant les deux premières années de sa mise en œuvre, c'est-à-dire au *kindergarten* (l'équivalent de la grande section de maternelle) et en *grade 1* (l'équivalent du cours préparatoire). Ils ne sont égaux qu'au tiers de ce coût à partir du *grade 2* (l'équivalent du CE1). Pour calculer les effets de la réduction de la taille des classes en école élémentaire sur les salaires à l'âge adulte, ils ont tout d'abord fait l'hypothèse, sur la base d'études préexistantes (cF. Krueger et Whitmore, 2001, Tableau 3, p. 8), que la scolarisation dans une classe de taille réduite en cycle élémentaire a fait croître de 0,13 point d'écart-type les résultats aux examens de fin d'étude secondaire (à l'âge de 18 ans), et ont **appliqué un résultat dû à Neal et Johnson (1996)²¹⁶ stipulant qu'une augmentation d'un écart-type des résultats scolaires en fin de cycle secondaire (à 18 ans) est associée à des salaires supérieurs de 20 % onze ans après**. Ils supposent finalement qu'en accord avec l'augmentation plausible de la productivité aux États-Unis au cours des quarante années à venir, les salaires réels croîtront en moyenne de 1 % chaque année. Les données du *Current Population Survey* (CPS) de 1999 leur servent à évaluer les salaires annuels perçus par les individus âgés de 18 à 65 ans.

L'ensemble de ces hypothèses ont permis à Krueger et Whitmore de calculer le taux de rendement interne r^* du dispositif considéré qui résout l'équation suivante :²¹⁷

$$\sum_{t=1}^4 \frac{c_t}{(1+r^*)^t} = \left[\sum_{t=14}^{61} \frac{\delta \beta y_t}{(1+r^*)^t} \right]$$

Dans cette équation, c_t représente le coût annuel de la réduction de la taille des classes pour chaque élève, y_t est le salaire annuel perçu durant l'année t , δ est l'effet de la

²¹⁶ Neal D. et Johnson W. (1996), "The Role of Premarket Factors in Black-White Wage Differences", *Journal of Political Economy*, Vol 104(5), p. 869-895.

²¹⁷ Le taux de rendement interne r^* est celui qui égalise les bénéfices et les coûts composant la formule de la VAN.

scolarisation dans une classe de taille réduite sur le score aux résultats scolaires en fin de cycle secondaire, et β est le paramètre qui mesure l'effet d'une augmentation d'un écart-type des résultats scolaires à la fin du cycle secondaire sur les salaires ultérieurs. Les bornes des sommes apparaissant dans cette formule correspondent respectivement aux âges 5 ($t = 1$), 8 ($t = 4$), 18 ($t = 14$) et 65 ($t = 61$).

Le taux de rendement interne r^* - taux qui annule la VAN - calculé par Krueger et Whitmore est égal à 5,2 %, 6,2 % ou 7,3 % selon que le taux annuel de croissance des salaires futurs (en termes réels) est prévu être nul, égal à 1 % ou à 2 %. Ces valeurs sont relativement faibles et permettent à Krueger et Whitmore de privilégier un scénario dans lequel le taux d'actualisation est égal à 4 %, soit le taux de rendement réel observé des titres d'État américains. Sous cette hypothèse, les bénéfices directs d'une réduction de la taille des classes seraient respectivement supérieurs de 43 %, de 100 % ou de 190 % aux coûts de la mesure dans le cas où le taux annuel de croissance de la productivité serait nul, égal à 1 % ou à 2 %. Ils le seraient encore plus si le taux d'actualisation était égal à 3 % : ils seraient alors deux, trois ou quatre fois supérieurs aux coûts lorsque le taux de croissance de la productivité passerait de zéro à 2 %. Dans cet exercice, les seuls cas dans lesquels les bénéfices deviennent inférieurs aux coûts sont ceux dans lesquels le taux d'actualisation est supérieur à 5 % et le taux anticipé de croissance de la productivité est nul.²¹⁸

Contrepoint : Retrouve-t-on dans les données ces rendements salariaux ?

Dans une étude plus récente, Chetty *et al.* (2011)²¹⁹ ont examiné la situation des enfants ayant participé au programme STAR plus de vingt après leur entrée dans l'enseignement primaire. C'est là une chose tout à fait remarquable, car il est rare de pouvoir connaître le devenir à long terme des bénéficiaires d'une politique publique, notamment éducative. Cette connaissance est pourtant essentielle parce qu'elle permet tout à la fois de tester réellement l'efficacité de la politique en question et de réviser si nécessaire les conclusions de l'analyse coûts-bénéfices qui a pu être faite au moment de sa mise en place, ou peu après. Pour réaliser un suivi à long terme, Chetty *et al.* ont réussi à apparier les données permettant d'identifier les enfants ayant participé au programme STAR aux données fiscales américaines. L'étude de Chetty *et al.* (2011) montre que les élèves qui ont été affectés à des classes de taille réduite accèdent plus fréquemment à l'enseignement supérieur. Toutefois, en raison d'une grande imprécision dans les estimations (liée à une forte variabilité et un nombre d'observation restreint en raison de la difficulté à retrouver dans les bases statistiques tous les individus ayant participé au programme), les auteurs ne trouvent pas de différence significative entre les salaires des individus des deux groupes (bénéficiaires et

²¹⁸ Bouguen, Grenet et Gurgand (2017) ont mis en œuvre un calcul du même type afin de réaliser une analyse coûts-bénéfices *ex ante* du dédoublement des classes de CP et de CE1 dans les établissements situés dans les réseaux d'éducation prioritaire renforcés (Rep+). Ce dédoublement a été mis en œuvre par le gouvernement français aux rentrées scolaires 2017 et 2018. Leur résultat principal est conforme à celui précédemment obtenu par Krueger et Whitmore, puisqu'ils trouvent « qu'une telle politique est rentable dès qu'elle permet d'augmenter le salaire futur de ses bénéficiaires de 1 %, *via* ses bénéfices sur la trajectoire scolaire ». Voir Bouguen A., Grenet J. et Gurgand M. (2017), « La taille des classes influence-t-elle la réussite scolaire ? », note n°28 de l'Institut des politiques publiques, Paris. Cette étude a été présentée devant le groupe de travail par Marc Gurgand.

²¹⁹ Chetty R., Friedman J., Hilger N., Saez E., Whitmore Schanzenbach D. et Yagan D. (2011), "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star", *The Quarterly Journal of Economics*, Vol. 126(4), p. 1593-1660.

non-bénéficiaires) à l'âge de 27 ans. Les intervalles de confiances autour des gains salariaux sont majoritairement positifs mais ne permettent pas d'exclure un effet nul à un seuil de 95 %.

Conclusion et ouverture

La façon dont Krueger et Whitmore (2001) ont conduit leur calcul a plusieurs mérites : 1) ils n'ont pas proposé un chiffrage unique de la différence entre coûts et bénéfices directs de la mesure considérée (*i.e.*, une réduction de la taille des classes à l'école élémentaire), ils ont fait dépendre celle-ci de plusieurs valeurs plausibles du taux de croissance de la productivité, et donc des salaires, et de la valeur du taux d'actualisation ; 2) ils ont retenu les valeurs du taux d'actualisation à privilégier en fondant celles-ci sur la valeur du taux de rendement interne estimé et des taux de rendement observés des bons d'État américains, et 3) ils ont identifié les cas dans lesquels les coûts deviennent supérieurs aux bénéfices. Toutefois, ils n'ont pas proposé une estimation des bénéfices indirects de la mesure et des coûts évités qu'elle induit. Ils n'ont également pas tenu compte des effets d'équilibre général, notamment parce qu'ils ont négligé les conséquences que la mesure pourrait avoir sur l'offre et la qualité des nouveaux enseignants nécessaires à sa mise en œuvre.

Le cas de la Californie est de ce point de vue particulièrement instructif. L'état de Californie imposa par la loi en 1996 une réduction d'un tiers de la taille des classes du *kindergarten* jusqu'au *grade 3* sur le modèle du programme STAR. Cette loi eut pour conséquence directe le recrutement de milliers de nouveaux enseignants. Jepsen et Rivkin (2002, 2009)²²⁰ pointent les difficultés à laquelle les autorités ont dû rapidement faire face. Mécaniquement, les nouveaux enseignants recrutés à la rentrée scolaire 1997 avaient une très faible expérience professionnelle. Mais là n'était pas le plus grave : de nombreux enseignants préalablement affectés à des classes de quartiers défavorisés demandèrent leur mutation dans des classes de taille réduite situées dans des zones plus aisées. Ce fut notamment le cas d'enseignants de classes des *grades 4 et 5* (les équivalents du CM1 et du CM2 en France) qui accueillaient beaucoup d'enfants des minorités afro-américaines et hispaniques. De nombreux nouveaux enseignants peu expérimentés furent alors affectés à ces classes, dont la taille n'était pas modifiée par la loi. Il en résulta une baisse du niveau dans ces classes. Certes, la mesure eut un effet global positif sur le niveau des élèves : dans les classes de *grade 3* (CE2) dont la taille fut réduite de 10 élèves, la proportion d'élèves crédités d'un score supérieur au niveau médian national s'accrut de 4 % en mathématiques et de 3 % en lecture. Mais cela ne fut pas le cas dans les classes accueillant les nouveaux enseignants : dans celles-ci, souvent situées dans des quartiers défavorisés, la mesure n'eut pas d'effet. Les effets bénéfiques d'une réduction de la taille des classes peuvent donc être amoindris, voire annulés, lorsque la qualité des enseignants recrutés est insuffisante.

²²⁰ Jepsen C. et Rivkin S. (2002), "Class Size Reduction, Teacher Quality, and Academic Achievement in California Elementary Public Schools", San Francisco: Public Policy Institute of California ; Jepsen C. et Rivkin S. (2009), "Class Size Reduction and Student Achievement: The Potential Tradeoff between Teacher Quality and Class Size", *The Journal of Human Resources*, Vol. 44(1), p. 223-250.

Que retenir ?

- Évaluation socioéconomique ex-post de l'effet de la réduction de la taille des classes.
- Analyse coûts-bénéfices basée sur une expérimentation aléatoire contrôlée et un suivi longitudinal long.
- Valorisation monétaire des performances scolaires à partir de résultats de littérature
- Mise en œuvre claire d'un calcul de VAN-SE
- Discussion détaillée des hypothèses de calcul et de la sensibilité aux principaux paramètres
- Prise en compte uniquement des effets directs sur le taux de salaire

Annexe A-II

Les analyses coûts-bénéfices du Washington State Institute for Public Policy

Le cadre général

Le *Washington State Institute for Public Policy* (WSIPP ci-après) est une organisation non gouvernementale et apolitique fondée en 1983 dans l'État de Washington aux États-Unis. Son but est de produire des recherches qui se veulent utiles et non partisans à destination des parlementaires et gouverneurs de l'État. L'organisation travaille en lien avec le législateur, les fonctionnaires de l'État de Washington et les experts pour s'assurer que les études produites sont à la fois pertinentes pour les décideurs et validées par les chercheurs.

Constitué d'une équipe de 22 personnes et présidé par un directoire composé de 16 personnalités (élus, chercheurs, etc.), le WSIPP produit des analyses coûts-bénéfices des politiques actives sur le marché du travail, des politiques de lutte contre la délinquance, des politiques éducatives, de santé, d'accès à l'enseignement supérieur, etc., selon une méthodologie détaillée et identique pour toutes les politiques évaluées. Le WSIPP propose ainsi une méthodologie permettant de comparer les rendements estimés de différentes politiques. Par la mise à disposition sur son site internet²²¹ de l'ensemble des revues de littérature et des applications produisant les estimations, le WSIPP facilite la compréhension et l'utilisation de ses travaux. En clair, si toutes les estimations qu'il produit peuvent être soumises à discussion, elles ont au moins le mérite de reposer sur un corpus d'hypothèses identifiables et éventuellement modifiables.

De façon très simplifiée, le WSIPP procède en trois étapes :

²²¹ <http://www.wsipp.wa.gov/>

- en premier lieu, il recense et synthétise la littérature mesurant les impacts de la politique examinée, puis calcule à partir de ce matériau les coefficients utilisés pour la méta-analyse ;
- Les tailles d'effets précédemment calculées sont ensuite monétisées puis comparées aux coûts par le biais du calcul de la valeur sociale nette actualisée et du ratio bénéfices-coûts ;
- Le WSIPP estime enfin l'incertitude associée au calcul du rendement d'un investissement social à l'aide de simulations de Monte Carlo. L'objectif est ici d'estimer la probabilité que l'estimation de ce rendement soit positive.

La méta-analyse des effets estimés

Le WSIPP procède tout d'abord à une revue systématique de la littérature académique correspondant à une politique donnée. Ce préalable lui permet de réaliser ensuite une méta-analyse à partir des études qui satisfont certains critères d'inclusion. Les études les moins rigoureuses sont ainsi écartées et ne sont conservées que celles présentant un niveau de validité interne suffisant. Plus précisément, seules sont retenues les études fondant leurs résultats sur l'existence d'un groupe de contrôle (expérimental ou non) et employant au moins une technique économétrique permettant de corriger les biais de variables omises ou de causalité inverse. Le WSIPP ne retient que des études réalisées aux États-Unis ce qui limite les possibilités d'utiliser leurs travaux en dehors des USA.

Des ajustements sont ensuite réalisés pour tenir compte de la nature des méthodes statistiques employées et pour calculer la précision de chaque estimation (qui est mesurée par son erreur-type²²²). Les valeurs utilisées dans la suite de l'analyse sont donc une moyenne pondérée des effets obtenus dans plusieurs évaluations indépendantes d'une même politique.

Le WSIPP propose par ailleurs une façon de remédier au problème qui apparaît lorsqu'une étude fournit des résultats sur certains effets de la politique considérée (par exemple, ses effets sur le niveau d'éducation atteint), mais pas sur d'autres (par exemple, ses effets sur les salaires futurs), alors que la littérature académique montre que ces deux types d'effets sont statistiquement associés. Un exemple supplémentaire est celui d'un programme de soutien à la fonction parentale, qui se trouve avoir un effet sur le niveau des violences faites aux enfants mais pour lequel l'analyste n'observe pas de réussite scolaire, et ce alors que de nombreuses études ont mis en évidence un lien causal entre ces deux variables (maltraitance et réussite scolaire). Dans ces cas, **le WSIPP adopte une méthode permettant de « convertir » ces effets en une seule valeur monétisée. Pour ce faire, il construit une matrice de coefficients méta-analytiques reliant des résultats intermédiaires à d'autres variables qui ne sont généralement pas observées dans les enquêtes de suivi. Cette méthode lui permet de réaliser une estimation des effets potentiels de long terme (non observés) à partir des effets observés sur un horizon plus court.**

²²² Par définition, l'erreur-type (ou erreur standard) d'une méthode d'estimation est l'écart-type de l'estimateur utilisé. Il fournit une mesure de la dispersion ou de la précision de cet estimateur.

La question des doubles comptes est un problème presque opposé : si certains résultats mesurés par plusieurs variables ont la même cause directe, à savoir le programme évalué, et s'ils sont simultanément monétisés dans le calcul des bénéfices, l'estimation risque de surestimer, ou au moins mésestimer, le véritable rendement social du programme. Par exemple, une amélioration des résultats scolaires en fin de second cycle et une plus grande probabilité d'avoir un diplôme de l'enseignement supérieur sont interprétées par le WSIPP comme la conséquence d'une même cause, à savoir une augmentation des capacités cognitives induites par le programme (par exemple, une réduction de la taille des classes à l'école élémentaire). La question des doubles comptes est à la fois un problème de mesure (plusieurs variables représentent la réussite scolaire) et un problème plus théorique : qu'est-ce qui est le reflet d'une augmentation du capital humain et qu'est-ce qui dépend d'autres facteurs.

Le WSIPP considère ainsi que le taux de diplomation dans l'enseignement secondaire, le nombre d'années d'études, ou encore la baisse des troubles mentaux et psychiques, sont différentes mesures d'un même phénomène, i. e. l'accumulation de capital humain qui a pour conséquence une meilleure insertion sur le marché du travail. De façon similaire, la réduction des montants de soins résultant d'une moindre consommation de stupéfiants ou d'une diminution des troubles mentaux et psychiques peut être aussi la conséquence de l'accroissement du niveau d'éducation et de capital humain.

Dans ce cas, **le WSIPP ne valorise que l'effet de plus grande amplitude** (i.e. de plus grande valeur absolue) **lorsque les effets sont tous positifs ou tous négatifs. Lorsque certains effets sont positifs et d'autres négatifs, il retient seulement l'effet positif de plus grande taille et l'effet négatif de plus grande valeur absolue.** Si le risque de doubles comptes concerne des effets liés, seul le résultat le plus « tardif » (i.e. éloigné de l'intervention initiale) est conservé.

Pour l'évaluation de l'effet d'une réduction de la taille des classes, le WSIPP exploite les résultats de 55 études et retient 77 estimations pour calculer la statistique de taille d'effet utilisé. Cet effet est interprété comme un effet marginal et le calcul suppose que l'effet moyen est linéaire.

Ainsi, lorsque la taille de la classe diminue d'un élève, les résultats scolaires de ceux restant dans la classe augmentent environ de 1,1 % d'écart type à 17 ans. Ce résultat est comparable à celui obtenu dans l'étude de Krueger et Whitmore (2001) qui a été présentée dans la sous-section précédente, et qui est incluse dans la liste des études retenues par le WSIPP pour sa méta-analyse : selon Krueger et Whitmore (2001), une réduction de 10 élèves correspond à un gain de 13 % d'écart type en CE2, ce qui est similaire au résultat du WSIPP sous l'hypothèse de linéarité de l'effet moyen.

Tableau 2 – Méta-analyse de l'effet d'une réduction marginale de la taille de la classe en dernière année de maternelle (un élève en moins dans chaque classe)

Variable de résultat	Âge au moment de l'intervention	Nombre d'estimations	Nombre d'observations dans le groupe traité	Tailles d'effet ajustées et erreurs standards utilisées dans l'analyse coûts-bénéfices						Taille d'effet non ajusté	
				1 ^{re} estimation			2 ^e estimation				
				ES	SE	Âge	ES	SE	Âge	ES	p-value
Diplôme du secondaire	5	77	1000	0.015	0.005	5	0.015	0.005	17	0.015	0.005
Résultats scolaires	5	77	1000	0.036	0.013	5	0.011	0.005	17	0.036	0.005

Remarques : ES est l'abréviation d'estimation, SE celle de Standard Error (erreur standard) ; la p-value (ou p-valeur) est comparée à un seuil préalablement défini (traditionnellement 5 %). Si la p-valeur est inférieure à ce seuil, le résultat du test est déclaré « statistiquement significatif ». Dans le cas contraire, si la p-valeur est supérieure au seuil, on ne peut rien conclure quant aux hypothèses formulées.

Source : <http://www.wsipp.wa.gov/BenefitCost/Program/204>

La monétisation des effets et le calcul du rendement social

À partir des méta-analyses permettant de mesurer les effets du programme considéré sur différentes variables d'intérêt, le WSIPP attribue une valeur monétaire à toute variation d'une taille d'effet selon des méthodologies décrites très précisément dans le guide technique. Chaque valorisation est obtenue à l'aide d'une méthode spécifique fondée sur une analyse de données précises qu'il ne nous est pas possible de résumer en détail. On peut toutefois indiquer deux exemples :

- les rendements salariaux de l'éducation sont obtenus à partir de régressions de Mincer estimées à partir du *Current Population Survey*²²³ ;
- le coût des délits comprend le coût subi par les victimes et le coût pour les finances publiques (coût d'incarcération, coût d'un jugement, etc.), tels qu'ils sont estimés à partir de données fournies par le *United States Department of Justice*.

Plusieurs ajustements sont également réalisés dans le calcul, notamment ceux relatifs à la valeur statistique de la vie humaine et à la « perte sèche » (*deadweight loss*) pour les finances publiques. La valeur de la vie humaine est fondée sur une estimation déduite des recherches académiques ; elle est généralement obtenue par évaluation contingente (environ 7 millions de dollars 2017), montant dont sont soustraites les variations des soins de santé et de sécurité sociale, les variations de revenus et de production domestique.

La « perte sèche » pour l'économie reflète la perte de surplus social lorsqu'un dollar d'impôt est utilisé pour financer un programme, ou le gain lorsqu'un programme réduit les contributions fiscales. L'idée générale est que l'intervention de la puissance publique financé par impôt modifie l'équilibre concurrentiel et aboutit, dans un modèle classique, à une perte sèche pour la société. Cet effet d'équilibre est difficile à appréhender sans modèle structurel. C'est pourquoi le WSIPP calcule la perte sèche à partir du gain ou de la perte nette pour les

²²³ Le *Current Population Survey* (CPS) est une enquête mensuelle réalisée par le United States Census Bureau pour le Bureau of Labor Statistics (BLS) auprès d'environ 60 000 ménages américains.

finances publique et utilise une valeur modale de 50 % et deux bornes (égales à 0 % et 100 %) pour les simulations tirées au hasard dans une distribution suivant une loi triangulaire. Ainsi, si un programme financé par l'impôt coûte 1 000\$ par participant et s'il permet de réduire de 600\$ les coûts pour les finances publiques en raison d'une réduction des frais judiciaires ou d'une augmentation des recettes fiscales résultant de revenus plus élevés, le coût d'opportunité pour les finances publiques s'élève à une perte sèche de 200 dollars avec une valeur modale fixée à 50 % (des 400 \$ de coûts nets d'économies). En pratique dans le modèle le calcul est effectué à chaque année et actualisé.

Selon le WSIPP, la valorisation des effets d'une réduction de la taille des classes de maternelle d'un élève :

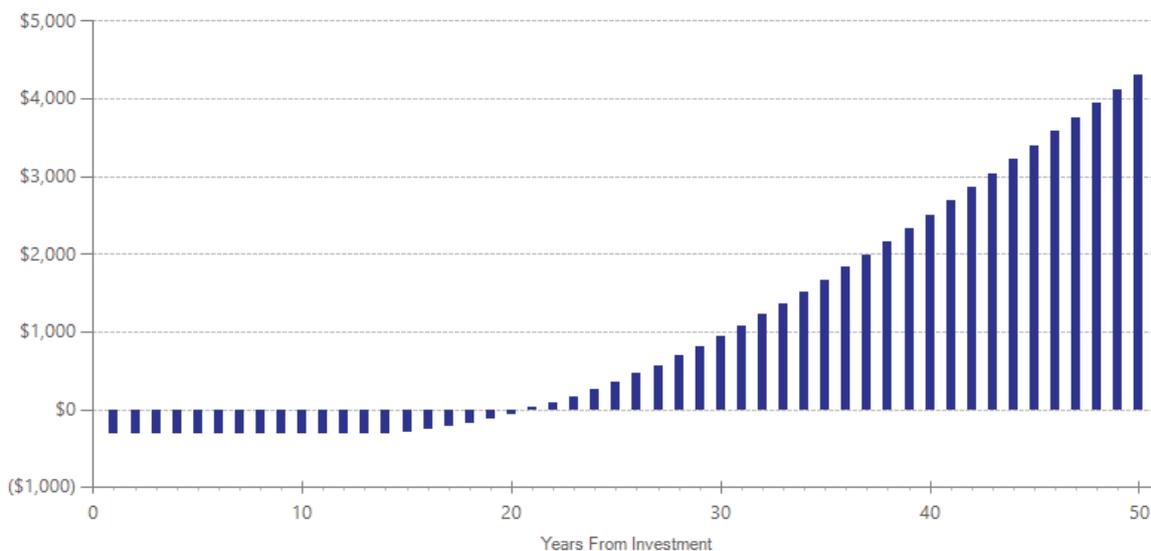
- conduirait à économiser sur cinquante ans 14 dollars (de 2017) correspondant à une réduction de la délinquance et 18 dollars correspondant à une réduction des soins de santé ;
- rapporterait 1 745 dollars correspondant à une augmentation des revenus du travail ;
- impliquerait pour la collectivité une perte sèche de 107 dollars.

La somme des effets monétisés actualisés par individu scolarisé en maternelle dans une classe réduite d'un élève serait donc de 1 670 dollars de 2017 (cf. dernière colonne du Tableau 5). Sachant que cette mesure coûte, d'après le WSIPP, 214 dollars (de 2017) en moyenne par élève, sa valeur actuelle nette est estimée à 1 456 dollars, soit un ratio bénéfices/coûts de $1\,670/214 = 7,81$. Cette valeur est une moyenne annuelle calculée sur les cinquante années suivant l'intervention. La figure 7 indique comment le bénéfice net cumulé, année après année, évolue au cours de ces 50 années. Ce bénéfice net cumulé ne devient positif qu'au bout de 25 années, et croît régulièrement par la suite pour atteindre 5 000 dollars (non actualisés) au terme de ces cinquante années. Cet exemple illustre clairement le fait que pour juger les politiques d'investissement social, il faut pouvoir estimer leur rendement social sur le long terme.

Tableau 3 – Bénéfices associés à une réduction marginale de la taille d'une classe de maternelle (un élève en moins par classe)

Bénéfices liés à une variation :	Bénéfices pour les...				
	Contribuables	Participants	Autres	Indirect	Total
De la délinquance	\$4	\$0	\$8	\$2	\$14
Des revenus du travail résultant d'un plus haut niveau de diplôme	\$418	\$920	\$407	\$0	\$1,745
Des frais des soins de santé résultant d'une meilleure éducation	\$120	(\$33)	(\$130)	\$61	\$18
D'un ajustement de la perte sèche pour l'économie (<i>deadweight loss</i>)	\$0	\$0	\$0	(\$107)	(\$107)
Total	\$542	\$887	\$286	(\$45)	\$1,670

Figure 7 – Bénéfices nets cumulés d'une réduction marginale de la taille d'une classe de maternelle (un élève de moins par classe)



Sources : <http://www.wsipp.wa.gov/BenefitCost/Program/204>

Les analyses de sensibilité et la méthode de Monte Carlo

L'ACB qui vient d'être présentée repose sur de nombreuses hypothèses et paramètres estimés. Si les hypothèses et les valeurs calibrées sont modifiées, les résultats le sont aussi. Il est donc important de savoir dans quelle mesure les résultats de l'ACB dépendent des hypothèses faites et des valeurs retenues.

Pour ce faire, le WSIPP utilise des simulations de Monte Carlo qui consistent à réestimer le ratio bénéfices/coûts ou la valeur actualisée nette des centaines, voire des milliers, de fois en faisant varier les paramètres du modèle. La méthode consiste à tirer les nouveaux paramètres au hasard dans les distributions statistiques des effets estimés. Ces distributions sont supposées être des lois normales dont la moyenne est la valeur de l'effet estimé et l'écart-type celui de l'effet estimé. Cette méthode est une variante des simulations par méthode de Monte-Carlo. Les coûts et le taux d'actualisation sont échantillonnés à partir de lois triangulaires. En général, le WSIPP réalise 10 000 tirages.

Une fois cet exercice terminé, le WSIPP calcule la moyenne des VAN simulées. Il en déduit une mesure d'incertitude qui est égale à la proportion de simulations pour lesquelles la VAN est positive. Dans l'exemple de la réduction de la taille des classes, 98 % des simulations correspondent à un rendement net positif. Mais de nombreuses autres politiques aboutissent à des proportions beaucoup plus faibles de VAN positives, ce qui signifie que la méthode ne permet pas de conclure avec une confiance suffisante à l'efficacité du programme examiné. C'est le cas des programmes de « *summer books* » qui consistent à envoyer chaque semaine à des enfants des livres à lire pendant les vacances. L'analyse coûts-bénéfices menée à partir de la méta-analyse de trois évaluations indépendantes aboutit à un ratio bénéfices/coûts égal en moyenne à 24 (soit un bénéfice de 24 dollars par dollar investi), mais la probabilité que ce rendement soit positif est seulement de 56 %, soit à peine plus d'une chance sur deux.

Que retenir ?

- Évaluations *ex ante* de nombreuses politiques d'investissement social
- Démarche systématique, transparente et modèle identique pour toutes les politiques
- Revue systématique de littérature et méta-analyse pour recenser les effets
 - Critères d'inclusion trop ouverts pour ce qui est de la méthode, et trop fermés pour ce qui est du zonage géographique (USA seulement).
- Revue systématique de littérature et méta-analyse pour recenser les variables médiatrices et leur attribuer un coefficient de « conversion » d'effets
- Prise en compte du risque de double comptage avec une règle de décision préalable
- Valorisation monétaire de tous les effets recensés
- Projection des valeurs sur tout le cycle de vie.
- Analyse de sensibilité par simulation

Annexe A-III

Comment choisir la prévention en santé la plus efficiente ?

L'exemple de la modélisation sur la population australienne²²⁴

Cadre général

L'Américain Christopher Murray et l'Australien Alan Lopez, alors ensemble à l'Université d'Harvard, ont proposé un indicateur synthétique de santé, nommé le DALY²²⁵. Ce dernier représente les années de vie en bonne santé perdues par rapport à une situation idéale dans laquelle les hommes atteindraient l'âge de 80 ans en bonne santé et les femmes 82,5 ans. Cet indicateur a l'avantage de condenser en un seul chiffre les années de vie perdues pour décès prématuré et les années vécues avec une incapacité liée à un problème de santé²²⁶. Le niveau de cette incapacité est maintenant établi à partir d'enquête en population générale. Il est alors demandé aux personnes interviewées de déterminer à partir de « vignettes » laquelle de deux personnes présentées est en meilleure santé²²⁷.

²²⁴ Cette sous-section a été rédigée par Pierre Arwidson, directeur adjoint de la prévention et de la promotion de la santé, Santé Publique France.

²²⁵ Daly : disability adjusted life year (année de vie corrigée des incapacités ou AVCI)

²²⁶ $Daly = YLL$ (years of life lost due to premature death) + YLD (years lived with disability)

²²⁷ Par exemple, une vignette présente une personne capable de ne marcher que sur de courtes distances, se sentant bien au repos mais évitant de faire des efforts importants car elle s'essouffle et se sent vite fatiguée. Une autre personne est décrite comme toujours essoufflée, et se fatiguant vite, même au repos. Elle évite toute activité physique de peur que ses difficultés à respirer s'aggravent. Voir Salomon JA, Haagsma JA, Davis A, de Noordhout CM, Polinder S, Havelaar AH, et al. Disability weights for the Global Burden of Disease 2013 study. *Lancet Glob Health*. 2015;3(11):e712-23.

Leur objectif était de disposer d'un indicateur simple qui permette de comparer les poids de morbi-mortalité associés aux différents problèmes de santé. Cet outil permet, en théorie, une répartition équitable des ressources entre ces différents problèmes en fonction du poids que l'on accorde à leur morbidité. En général, cette répartition dépend malheureusement, que ce soit au niveau national ou international, des capacités d'influence des différents groupes de pression souhaitant agir sur chacun des problèmes de santé.

Murray et Lopez ont ensuite rejoint l'Organisation mondiale de la santé et ont appliqué leur concept à la mesure de la santé mondiale. Le rapport mondial de la santé de 1999²²⁸ a été le premier à proposer cette mesure synthétique de la santé pour tous les pays. Depuis, des mises à jour de la charge mondiale de morbidité (*global burden of disease study*²²⁹) sont régulièrement effectuées avec des améliorations à chaque vague, la dernière estimation ayant été faite pour l'année 2016²³⁰.

L'enjeu suivant était de comparer les différents moyens de prévention aptes à réduire la charge de morbidité de la façon la plus efficiente. C'est ce à quoi se sont attachés deux équipes universitaires australiennes, l'une localisée à Melbourne, dans laquelle travaille Alan Lopez, et l'autre de l'université du Queensland. Ces équipes avaient acquis une expérience sur des travaux antérieurs sur le cancer, la santé mentale, les maladies cardiaques et l'obésité. Leur objectif était à nouveau d'aider à la décision afin de pouvoir choisir la meilleure allocation possible des ressources financières.

Démarche générale

Leur projet consistait à **évaluer le rapport coût-efficacité de 100 interventions préventives des maladies non transmissibles** et des facteurs de risque associés afin d'aider à déterminer celles étant les plus efficientes. Le projet a inclus également l'analyse de 50 traitements de diverses maladies non transmissibles et des moyens de lutte contre certaines maladies infectieuses. Le projet de recherche a duré cinq ans et le rapport a été publié en 2010²³¹.

Le cadre de l'analyse économique était limité aux dépenses de santé dans le contexte australien. La modélisation, qui est comparable à une microsimulation dynamique²³², a été appliquée à la **population australienne de 2003 que les auteurs ont fait vieillir de quatre-vingts ans sans ajout de naissances**. Des analyses de sensibilité ont été réalisées, en faisant varier systématiquement les paramètres aléatoirement au sein de leurs marges d'incertitude pour vérifier la solidité des résultats. Un taux d'actualisation de 3 % a été appliqué. Il s'agissait de comparer les pratiques préventives et de soins mises en place en 2003 ainsi que leurs coûts.

²²⁸ World Health Organization (2000): The world health report 1999— making a difference. Genève: WHO.

²²⁹ Murray C. et Lopez A. (1997), "Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study", *Lancet*, 349, p. 1436-1442.

²³⁰ GBD Risk Factors Collaborators (2016), "Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015", *Lancet*, 388, p. 1659-724.

²³¹ Vos T., Carter R., Mihalopoulos C., Veerman L., Magnus A., Cobiac L. *et al.* (2010), *Assessing Cost-Effectiveness in Prevention*. Brisbane, Melbourne: University of Queensland and Deakin University, 101 pages.

²³² Dans ce domaine, comme dans d'autres, on ne peut faire l'économie d'évaluations *ex ante* par simulation, et il est difficile d'éviter que celles-ci soient réalisées au niveau micro-économique.

L'introduction dans le modèle des impacts bénéfiques des interventions a bien sûr comme limite la qualité des études réalisées, leur transférabilité à un contexte spécifique et la capacité à les disséminer. Ces facteurs ont été pris en compte *via* des groupes d'experts *ad hoc*. Par exemple, la qualité des données sur la prévention de l'obésité n'est pas bonne, alors qu'elle l'est davantage sur la prévention dans le domaine de la consommation de tabac et d'alcool. Cela a permis de classer chaque intervention selon deux axes, d'une part leur capacité à éviter la perte d'années de vie en bonne santé (DALY), autrement dit leur efficacité, et d'autre part leur efficience.

Les 150 interventions de prévention ont été classées en cinq catégories selon qu'elles sont :

- dominantes : elles améliorent la santé et permettent de faire des économies,
- efficaces et très efficaces : elles améliorent la santé et coûtent moins de 10 000 dollars australiens par année de vie en bonne santé (DALY)
- efficaces et efficaces : elles améliorent la santé et coûtent entre 10 et 50 000 dollars australiens par DALY
- efficaces et inefficaces : elles améliorent la santé mais coûtent plus de 50 000 dollars australiens par DALY
- dominées : elles n'améliorent pas la santé pour un coût donné, ou il existe une alternative plus efficace.

La liste des interventions dominantes est présentée dans le tableau 6.

L'application de l'ensemble des mesures dominantes présentées dans le tableau 6 permettrait de gagner un million de DALYs sur l'ensemble de la vie de la population australienne de 2003 (la surface sous la courbe verte sur la figure 8). Ces mesures coûteraient 4,6 milliards de dollars australiens (la somme des barres rouges de la même figure), mais pourraient permettre d'économiser 11 milliards de dépenses de santé (la somme des barres bleues). Quatorze pour cent de la dépense devraient être investis la première année. Une faible augmentation des coûts de santé serait à prévoir à partir de 2059 en raison des gains d'espérance de vie obtenus.

Après ce travail analytique, les auteurs proposent à l'aide d'une modélisation la meilleure **combinaison de stratégies de prévention dans de nombreux domaines**, comme la prévention de la consommation d'alcool²³³, la prévention des risques solaires²³⁴, la promotion de l'activité physique²³⁵ ou la consommation de fruits et légumes²³⁶. Par exemple, dans le domaine de la prévention des dommages liés à l'alcool, 410 millions de dollars australiens et 118 000 DALYs seraient économisés sur une période de quatre-vingts ans si l'on appliquait en Australie une combinaison de mesures telles que des taxes importantes sur la

²³³ Cobiac L., Vos T., Doran C. et Wallace A. (2009), "Cost-effectiveness of interventions to prevent alcohol-related disease and injury in Australia", *Addiction*, Vol. 104(10), p. 1646-55.

²³⁴ Shih S., Carter R., Sinclair C., Mihalopoulos C. et Vos T. (2009), "Economic evaluation of skin cancer prevention in Australia". *Preventive Medicine*, Vol. 49(5), p. 449-53.

²³⁵ Cobiac L., Vos T. et Barendregt J. J. (2009), "Cost-effectiveness of interventions to promote physical activity: a modelling study", *PLoS Medicine*, Vol. 6(7), e1000110.

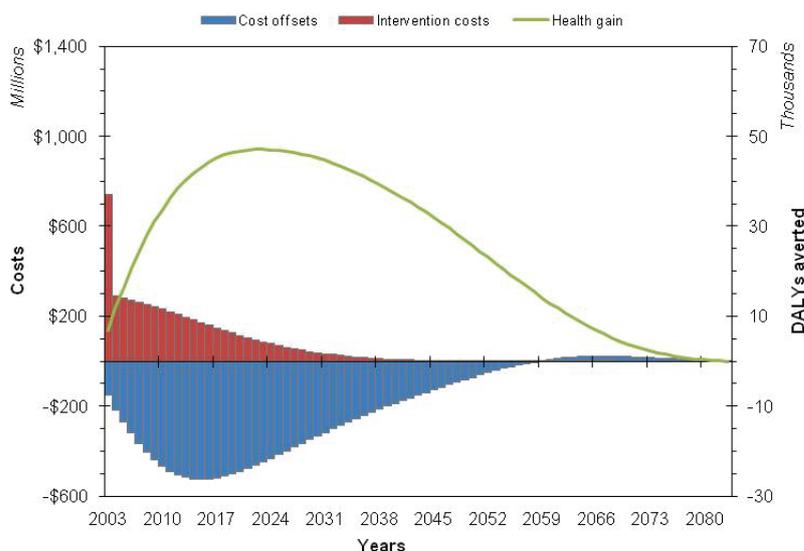
²³⁶ Cobiac L., Vos T. et Veerman J. L. (2010), "Cost-effectiveness of interventions to promote fruit and vegetable consumption", *PloS one*, Vol. 5(11), e14148.

consommation d'alcool, l'interdiction totale de publicité, l'augmentation de l'âge légal de consommation à 21 ans, la généralisation du repérage précoce et l'intervention brève dans le système de soins, la réduction de l'accès au produit (horaires d'accès réduits), des campagnes médiatiques contre l'alcool au volant, des tests aléatoires d'alcoolémie au volant et les traitements résidentiels.

**Tableau 4 – Liste des mesures dominantes
(qui permettent des gains de santé et un retour sur investissement)**

Thème	Mesure
Alcool	Taxe proportionnelle au degré d'alcool
	Accroissement de la taxe sur les boissons alcoolisées (+ 30 %)
	Interdiction de la publicité
	Accroître l'âge minimum de consommation d'alcool à 21 ans
Tabac	Accroître les taxes de 30 %
Activité physique	Promotion de l'utilisation de pedomètre
	Campagne médiatique de promotion de l'activité physique
Nutrition	Promotion au niveau local de la consommation de fruits et légumes
	Autolimitation de la quantité de sel dans les produits alimentaires transformés
	Réglementer la quantité de sel dans les produits alimentaires transformés
Masse corporelle	Taxe de 10 % sur les aliments sans intérêt nutritionnel
Pression artérielle et cholestérol	Programme local de santé cardiovasculaire
	Polypill pour les personnes ayant un risque supérieur à 5 %
Ostéoporose	Dépistage chez les femmes de plus 70 ans et mise sous traitement par alendronate
	Vaccination et immunoglobines pour les enfants nés de mères infectées ou à haut risque
	Vaccination des enfants à haut risque
	Vaccination sélective d'enfants de mères venant de pays avec forte endémie
Maladie rénale	Recherche de protéinurie chez les diabétiques et mise sous inhibiteurs de l'enzyme de conversion de l'angiotensine et inhibiteurs ACE pour les diabétiques
Santé mentale	Intervention après tentative de suicide
	Traitement pour les personnes à haut risque de psychose
Santé buccodentaire	Fluoration de l'eau de boisson

Figure 8 – Économies et gains de santé



Que retenir ?

- Évaluations *ex ante* par calculs coûts-efficacité en santé
- Microsimulation dynamique sur micro-données de politiques de santé validées par la recherche
- Validation par un comité d'expert des données probantes mobilisées
- Élaboration de « policy mix » par domaine à partir des résultats obtenus

Annexe A-IV

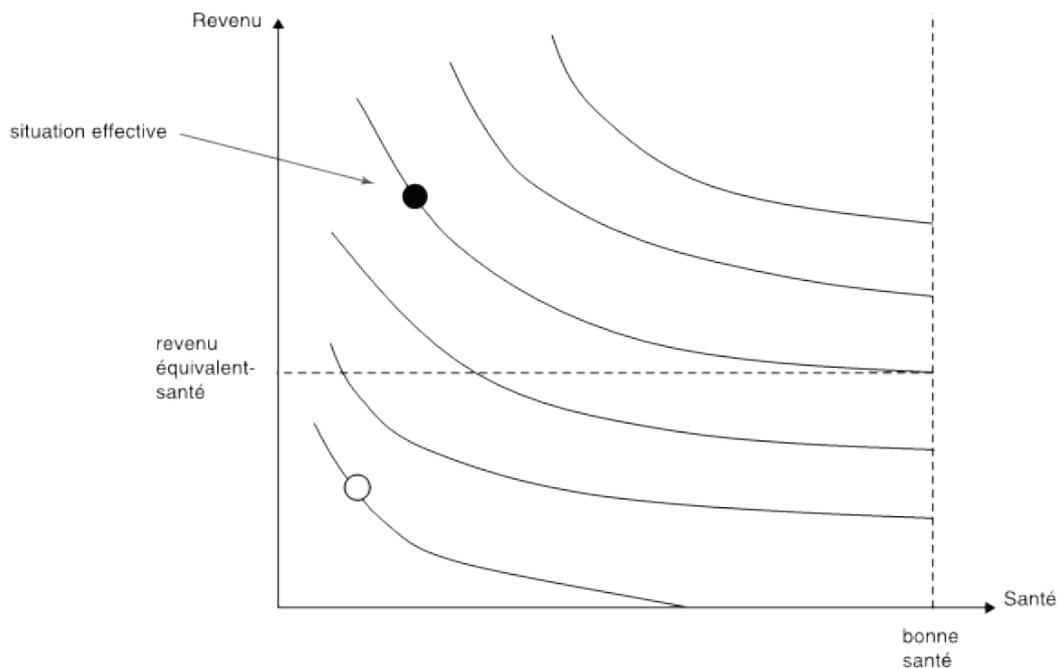
Prise en compte des enjeux d'équité dans l'analyse coûts-bénéfices : une application sur données françaises

Cadre général

Dans le champ de l'économie de la santé, les praticiens et chercheurs recourent avec réticence aux analyses coûts-bénéfices. Leur argument principal est que l'équivalent monétaire d'une amélioration de santé par le biais du consentement à payer est moralement très discutable, en particulier parce que les plus aisés peuvent avoir une disposition à payer plus élevée en raison de leur revenu. C'est donc l'approche coût-efficacité qui est ici privilégiée, à l'instar des indices *Quality Adjusted Life Years* (QALY) développé par le *National Institute for Health and Care Excellence* (NICE) et *Public Health England*. Mais l'utilisation des indices QALY ne règle pas pour autant les enjeux d'équité et de redistribution liés aux politiques de santé. Elle implique par ailleurs de disposer de valeurs tutélaires pour interpréter le résultat de l'évaluation.

L'approche proposée par Anne-Laure Samson et ses co-auteurs²³⁷ mobilise le concept de revenu équivalent-santé pour monétiser les gains en termes de santé. Ce concept correspond au revenu individuel effectif diminué du montant auquel la personne serait prête à renoncer pour être en parfaite santé. À niveau de revenu donné, le revenu équivalent-santé décroît quand la santé se détériore.

Figure 9 – Illustration du revenu équivalent santé, d'après Fleurbaey et al. ²³⁸ (2012)



Lecture : chaque ligne correspond à une courbe d'indifférence. L'individu repéré par le point noir est indifférent entre sa situation effective et la situation hypothétique où il serait en parfaite santé avec le niveau de revenu correspondant au trait pointillé. C'est ce niveau de revenu qui est qualifié de revenu équivalent santé et qui sert à caractériser le niveau de bien être commun à tous les individus de la courbe d'indifférence correspondante. Ce revenu équivalent santé reste toutefois indéfini pour l'individu repéré par le point blanc.

Contrairement à des indices d'utilité subjective, le revenu équivalent-santé a toutefois l'avantage de ne reposer que sur les préférences ordinales des individus, c'est-à-dire sur leurs préférences relatives et non absolues. En théorie, cette mesure devrait traduire les préférences des individus à l'égard de leur état de santé sans imposer d'hypothèses contraignantes sur leur fonction d'utilité, en particulier sans supposer que leurs préférences sont homogènes. S'il est possible de faire révéler correctement par un individu sa disposition à payer pour être en parfaite santé, le revenu équivalent-santé fournit alors une mesure permettant de **valoriser en unités de compte les gains de santé et de réaliser une analyse coûts-bénéfices**.

L'objectif est donc de comparer la valeur sociale de traitements en tenant compte des préférences individuelles en faveur d'une meilleure santé et donc en intégrant des considérations d'équité dans l'évaluation coûts-bénéfices.

²³⁷ Cette recherche a été présentée devant le groupe de travail par Anne-Laure Samson. Elle a donné lieu à la publication de l'article suivant : Samson A. L. et al. (2018), "Fairness in cost-benefit analysis: A methodology for health technology assessment", *Health Economics*, Vol. 27, p. 102-114.

²³⁸ Fleurbaey M., Luchini S., Schokkaert E. et Van de Voorde C. (2012), « Évaluation des politiques de santé : pour une prise en compte équitable des intérêts des populations », *Economie et Statistique*, n° 455-456, p. 11-36.

Samson *et al.* (2018)²³⁹ utilisent cette méthode pour calculer et comparer la valeur sociale de deux traitements contre l'hypertension. L'un de ces traitements apparaissait comme étant le moins cher selon les critères du modèle coût/efficacité utilisé par la Haute Autorité de Santé (HAS), le deuxième comme étant le plus efficace, un troisième servant de placebo. Cette étude a été réalisée en adoptant une approche *ex post* (les personnes sont interrogées par rapport à leur passé), permettant de déterminer à la fois la nature des préférences individuelles (puisque les personnes sont interrogées sur leur état de santé passé) et le type de modélisation mobilisée (voir *infra*). Par ailleurs, afin d'évaluer la sensibilité des résultats au type de questionnement proposé, une analyse *ex ante* était proposée aux individus.

Cette démarche est plus complexe que celle utilisée dans les analyses coûts-bénéfices présentées jusqu'alors ; en effet, outre la prise en compte des consentements à payer, les auteurs s'attachent à construire des trajectoires de santé où des chocs aléatoires peuvent survenir à chaque période. L'analyse consiste ainsi en plusieurs étapes :

1. Les auteurs estiment les consentements à payer pour différents profils à partir du revenu équivalent santé ;
2. Ils simulent ensuite l'ensemble des trajectoires de santé possibles pour les personnes souffrant d'hypertension ou non ;
3. Ils calculent les revenus des personnes et leur état de santé à chaque période, ainsi que les coûts liés aux parcours de soins pour chaque trajectoire simulée ;
4. À partir des résultats de la première étape, ils calculent les revenus équivalent santé correspondant à chaque état de santé à chaque date, et ce pour chaque trajectoire simulée ;
5. Ils agrègent ces résultats dans une fonction de bien-être social permettant d'accorder plus ou moins de poids aux personnes dont le revenu équivalent santé est le plus faible.

Estimation du revenu équivalent santé en France

Grâce à une enquête menée en 2009 auprès de 3 300 personnes interrogées en face à face, les auteurs de l'étude disposent d'informations sur l'état de santé des personnes (16 % souffrent d'hypertension), sur leurs caractéristiques sociodémographiques, leurs revenus et surtout sur une série d'éléments permettant de révéler leur disposition à payer pour être en parfaite santé.

Ces informations sont obtenues à partir de questions relatives à la façon dont les personnes interrogées évaluent (rétrospectivement) leur état de santé au cours des 12 derniers mois. On leur demande ensuite si elles eussent préféré éviter les problèmes de santé auxquels elles ont fait face mais en disposant d'un revenu plus faible, et si oui, quel eût été le montant maximal de la réduction correspondante de revenu. En moyenne, d'après cette enquête, les personnes sont prêtes à sacrifier 70 euros/mois pour être en parfaite santé, ce qui peut paraître faible. Les personnes souffrant d'hypertension ont une disposition à payer de 87 euros. Pour celles n'en souffrant pas, le montant équivalent serait de 62 euros.

²³⁹ Cf. *supra*.

La principale critique de ce type d'analyse porte sur la possibilité de faire révéler par les personnes enquêtées leur véritable disposition à payer, si tant est qu'elles la connaissent. Les questions posées sont souvent compliquées, elles peuvent parfois prêter à confusion. Il se peut en outre que certains sondés aient des comportements stratégiques lors de l'interrogation²⁴⁰. Ajoutons que, dans un pays comme la France où les frais de santé sont largement couverts par l'assurance maladie et où le reste à charge est faible, la disposition à payer révélée est difficilement interprétable. Ces limites ne sont toutefois pas spécifiques au problème ici traité, elles concernent la plupart des évaluations contingentes.

Modélisation des trajectoires de santé

Ces informations, en complément de celles fournies par un modèle coût-efficacité produit par *IMS Health* pour la Haute Autorité de Santé, sont ensuite utilisées pour simuler l'impact de différentes politiques publiques sur les revenus équivalent-santé des individus. Ce modèle coût-efficacité a été utilisé pour déterminer les probabilités d'occurrence de différents accidents cardio-vasculaires, selon que l'individu est ou non traité, mais aussi les probabilités que ces événements soient fatals, et ce en fonction du sexe, de l'âge, de la sujétion au diabète et de la consommation de tabac. Le modèle fournit également des informations sur les coûts des traitements et les coûts médicaux liés à l'occurrence d'une maladie. L'ensemble de ces données a permis de simuler pour chaque individu souffrant d'hypertension, toutes les trajectoires de santé possibles pendant les 10 prochaines années, et d'estimer, à caractéristiques socioéconomiques données, les conséquences de chaque traitement sur le revenu individuel au cours de ces 10 années. Les trajectoires possibles sont complexes : de nombreux événements peuvent survenir à chaque période, ces chocs ayant des conséquences sur les périodes suivantes. Pour représenter cette complexité, les auteurs utilisent un processus stochastique Markovien²⁴¹ pour faire survenir au cours de chaque période suivante un certain nombre d'événements (accident vasculaire cérébral, attaque cardiaque, etc.). Il est ensuite possible d'estimer les revenus équivalent-santé associés à chaque trajectoire et chaque traitement, année par année, et au terme des dix ans retenus, de les agréger dans une fonction de bien-être collectif. Les préférences concernant la santé et le revenu sont évaluées à partir de la situation actuelle des personnes et du revenu équivalent-santé fourni par l'enquête.

Pour limiter les difficultés calculatoires, les auteurs de l'étude précitée ont restreint le nombre d'occurrences possibles. Plus précisément, ils ont supposé que pour chaque individu, au plus un seul événement pouvait survenir chaque année, et qu'au total, pendant la période des dix années, ce nombre était limité à deux occurrences. Ces hypothèses engendraient 3 376 trajectoires pour chaque individu.

Les auteurs ont donc calculé pour chaque individu son état de santé, son revenu effectif et son revenu équivalent-santé, année après année, au sein de chaque trajectoire simulée. Ces estimations ont été ensuite agrégées sur la totalité de la période afin d'estimer une fonction de bien-être social. C'est à cette étape qu'est introduite la notion d'aversion à l'inégalité. La

²⁴⁰ Si la personne comprend que sa réponse peut avoir une influence sur ce qui lui est ensuite demandé, elle peut avoir intérêt à fournir par anticipation une estimation qui ne révèle pas ses vraies préférences. Par exemple, elle peut vouloir minimiser le revenu qu'elle se dit prêt à sacrifier si elle s'attend à ce que ses réponses affectent le remboursement des traitements.

²⁴¹ Il s'agit de chaînes de Markov.

fonction de bien-être social agrège les situations (trajectoires et revenus équivalent santé correspondants) et leur attribue plus ou moins de poids en fonction de ce paramètre d'aversion à l'inégalité. Dans ce cas précis, la fonction utilisée est dite à élasticité constante²⁴². Elle est de la forme :

$$SW = \frac{1}{1-\rho} \sum_p \sum_i \pi_{ip} \left(\sum_t Y_{ipt}^* \right)^{1-\rho}$$

Dans cette expression, l'indice p représente toutes les trajectoires possibles, π_{ip} est la probabilité d'une trajectoire p pour l'individu i (il s'agit de la fréquence d'apparition d'une trajectoire donnée parmi les 3376 possibles), Y_{ipt}^* est le revenu équivalent-santé de l'individu i au sein de la trajectoire p à la date t , et ρ est le paramètre d'aversion à l'inégalité. Dans cette fonction de bien-être social, la pondération que l'on accorde aux personnes les moins riches est d'autant plus forte que le paramètre ρ est élevé.

Résultats de l'évaluation

Les résultats de cette analyse montrent qu'au bout de dix ans, la stratégie la moins coûteuse est toujours préférée à la stratégie la plus efficace, quelle que soit la valeur du paramètre d'aversion à l'inégalité. Ce résultat s'explique par le fait que le coût des deux stratégies est de toute façon faible – même pour la stratégie la plus coûteuse – et réparti entre tous. Dans ce cas précis, l'analyse coûts-bénéfices ainsi réalisée produit les mêmes résultats qu'une analyse coûts-efficacité plus classique.

Cette étude permet également de montrer l'impact du choix de l'approche retenue, *ex ante* ou *ex post* : les analyses de sensibilité font en effet apparaître des résultats différents lorsqu'une approche *ex ante* est retenue. Le traitement « placebo » est alors préféré jusqu'à la 7^e année lorsque $\rho = 0$, et il est toujours préféré lorsque $\rho = 2$. Cette différence s'explique par le fait que, dans l'analyse *ex post*, les pondérations d'équité sont appliquées sur les revenus équivalents associés à toutes les trajectoires possibles, y compris les plus extrêmes, et non sur les espérances de revenu équivalent-santé, comme cela est fait dans l'analyse *ex ante*.

Au total, cette analyse révèle le potentiel de l'utilisation du revenu équivalent-santé : celui-ci permet de tenir compte simultanément des préférences individuelles et des enjeux d'équité dans la cadre d'une ACB. Ce type d'analyse peut être répliqué à de nombreuses politiques de santé et éventuellement à d'autres d'investissements sociaux, à condition de disposer à la fois de l'ensemble des informations nécessaires pour simuler les trajectoires possibles²⁴³ et d'enquêtes de qualité permettant d'observer non seulement l'état de santé et le revenu des individus mais aussi d'élucider leur disposition à payer pour être en bonne santé.

²⁴² Voir, par exemple, Atkinson A. (1970), "On the Measurement of Inequality", *Journal of Economic Theory*, Vol. 2, p. 244-263.

²⁴³ Ces informations sont tout autant nécessaire pour mener une analyse coûts-efficacité.

Que retenir ?

- Évaluations coûts-bénéfices des différents traitements dans la prise en charge de l'hypertension
- Analyse principale fondée sur une approche ex post, sachant qu'une approche ex ante est également réalisée de façon complémentaire
- Sortie du cadre habituel en santé autour de l'analyse coûts-efficacité
- Mobilisation d'une technique de préférences exprimées pour valoriser les bénéfices
- Simulation complexe de trajectoires de santé
- Possibilité de prendre en compte différents degrés d'aversion aux inégalités et de mesurer leur impact sur les résultats de l'évaluation
- Utilisation de ces valeurs pour construire des « pondérations éthiques » :
 - la prise en compte de l'éthique ne change pas les résultats dans ce cas thérapeutique précis
 - complexité de mise en œuvre
 - limite liée à l'usage d'évaluations contingentes.

Annexe A-V

Le programme *Perry Preschool*

Le programme et ses principaux résultats

L'une des expérimentations éducatives les plus commentées, notamment parce qu'elle a donné lieu à un suivi sur le très long terme, est le Perry Preschool Program.²⁴⁴ Ce programme a été mis en place entre 1962 et 1967 à l'initiative du psychologue David Weikart dans le district d'Ypsilanti (Michigan). L'évaluation a pu être réalisée grâce à une expérimentation conduite auprès de 123 enfants nés dans des familles très pauvres, essentiellement afro-américaines. Ces enfants, âgés de 3 à 4 ans, avaient préalablement obtenu de faibles résultats aux tests de QI ; 58 d'entre eux ont été aléatoirement affectés au groupe de traitement (ou groupe cible), les 65 autres ont formé le groupe de contrôle (ou groupe témoin). Les enfants du groupe cible bénéficiaient chaque jour de la semaine d'une demi-journée consacrée à leur éveil et aux apprentissages cognitifs, et ce pendant 8 mois durant l'année scolaire. L'encadrement était assuré par des éducateurs hautement qualifiés, formés aux pratiques pédagogiques prônées par David Weikart et inspirées des travaux pionniers de Jean Piaget, John Dewey et Lev Vygotsky. Ainsi, les enfants étaient encouragés à planifier et réaliser leurs activités par le biais d'un

²⁴⁴ Les informations relatives à ce programme sont disponibles sur les sites suivants :

<https://eric.ed.gov/?id=ED204426>, <https://eric.ed.gov/?id=ED366433>,
<https://highscope.org/fr/perry-preschool-study>.

processus de type « planification / exécution / révision ». Les éducateurs les incitaient également à faire des choix, à résoudre des problèmes, et à participer à des activités. Au lieu de donner des leçons, ils mettaient l'accent sur des questions de réflexion, telles que « Que s'est-il passé ? Comment as-tu fait ça ? Peux-tu me montrer ? Peux-tu aider un autre enfant ? ».

Le taux d'encadrement était particulièrement élevé, un éducateur s'occupant en moyenne de six enfants. En outre, chaque éducateur devait rendre visite à la famille de l'enfant une fois par semaine, environ pendant 90 minutes, afin d'évoquer avec les parents les progrès et les difficultés de l'enfant, la façon dont ils pouvaient l'accompagner dans le cadre du programme, mais aussi observer l'environnement familial de l'enfant.

Dès l'entrée à l'école élémentaire, les résultats scolaires des enfants du groupe cible étaient de loin supérieurs aux résultats des enfants du groupe de contrôle. Par la suite, les enfants du groupe cible ont été **plus longtemps scolarisés** (11,9 contre 11 années), ils ont **passé moins de temps dans les établissements spécialisés, réservés aux enfants présentant des déficiences mentales ou émotionnelles, des difficultés d'élocution ou d'apprentissage** (3,9 contre 5,2 années), ils ont **plus souvent obtenu un diplôme de l'enseignement supérieur** (66 contre 44 %). Dans le groupe cible, la proportion d'adolescentes enceintes était inférieure (0,6 contre 1,2 grossesse en moyenne). À 40 ans, les bénéficiaires du programme *Perry Preschool* ont été **moins souvent emprisonnés** (28 contre 52 %), ils percevaient un **revenu mensuel médian supérieur de 42 % à celui des membres du groupe de contrôle** (1 856 contre 1 308 dollars) et ils étaient moins nombreux à avoir bénéficié de l'aide sociale au cours des dix années précédentes (59 contre 80 %). **Toutefois, les critiques soulignent la petite taille de l'échantillon expérimental (123 enfants au total)**, l'absence d'effet substantiel à long terme sur le quotient intellectuel, et l'absence de significativité statistique de nombre de ses effets.

Les analyses coûts-bénéfices du programme

Dans les premières études proposant des ACB de ce programme, les estimations de son taux de rendement social étaient particulièrement élevées : il était de 16 % pour Rolnick et Grunewald (2003), de 17 % pour Belfield *et al.* (2006).²⁴⁵ Malheureusement ces deux études ne reportent pas les écarts-types des estimations. Par ailleurs, la première ne contient pas d'analyse de sensibilité, la seconde en propose une mais qui est trop concise. En effet, elle se limite à tester quelques valeurs de paramètres comme le taux d'actualisation mais ne tient pas compte par exemple de la précision des tailles d'effet estimées.

Indépendamment de cela, ces ACB du programme *Perry Preschool* tiennent très imparfaitement compte des quatre difficultés suivantes : a) le protocole de la randomisation a été compromis par le fait que certains enfants ont été réaffectés à l'un des deux groupes en dépit des résultats du tirage aléatoire préalablement réalisé ; b) les

²⁴⁵ Rolnick A. et Grunewald R. (2003), "Early childhood development: Economic development with a high public return", Technical Report, Federal Reserve Bank of Minneapolis, Minneapolis, MN.

Belfield C. R., Nores M., Barnett W. S. et Schweinhart L. (2006), "The High/Scope Perry Preschool Program: Cost-benefit analysis using data from the age-40 follow-up", *Journal of Human Resources*, Vol. 41(1), p. 162-190.

individus ayant participé au programme ont été suivis jusqu'à l'âge de 40 ans, mais pas au-delà ; si l'on veut obtenir une estimation des bénéfices et des coûts du programme sur une période plus longue, par exemple jusqu'à 65 ans, il est nécessaire de recourir à des méthodes d'appariement avec d'autres sources statistiques ; c) pour quelques individus n'ayant pas encore atteint 40 ans (*i. e.*, avant la fin de la période d'observation), certaines informations sont manquantes; et d) il est difficile d'attribuer des valeurs monétaires fiables à certaines variables telles que les délits, qui peuvent être plus ou moins graves.

Dans un article publié en 2010,²⁴⁶ Heckman et ses coauteurs ont proposé des solutions à chacun de ces problèmes et ont ensuite réestimé le taux de rendement annuel moyen du programme Perry Preschool qui, selon eux, serait compris entre 7 % et 10 %. La façon dont ils ont choisi d'attribuer une valeur monétaire aux délits et à leur jugement, qui est en soi un exercice difficile, explique presque entièrement la différence entre leur estimation du taux de rendement du programme et les précédentes, très supérieures. Par ailleurs, Heckman et al. (2010) ont calculé les écarts-types de la plupart des effets estimés du programme. Lorsque cela ne leur est pas possible, ils ont analysé la sensibilité de leurs résultats à certains paramètres qui ne peuvent pas être estimés, tels que le taux d'actualisation (qu'ils font varier de 0 % à 7 %), la « perte sèche » fiscale (« deadweight loss ») qui permet de financer le programme, ou encore la valeur monétaire d'une vie humaine (qui intervient dans le calcul du coût d'un meurtre). Heckman et al. (2010) ont utilisé plusieurs méthodes d'imputation pour compléter les informations manquantes (avant l'âge de 40 ans) ou celles qui ne peuvent être observées (entre 40 et 65 ans). Ils ont utilisé autant que possible les données officielles relatives aux coûts de l'éducation, de la délinquance ou des dispositifs d'aide sociale. Ils ont enfin reporté une analyse de sensibilité très complète qui permet de se faire une idée de la précision de leurs résultats. Comme dans les analyses antérieures, Heckman et al. (2010) ont évalué le coût initial (non actualisé) du programme à 17 759\$ (dollars de 2006) par enfant. Ce coût inclut les frais de fonctionnement (salaires des éducateurs et frais administratifs) et les coûts en matériel (mobilier, livres, etc.).

Estimations des « bénéfices négatifs » du programme (ou coûts sociaux)

Le programme semble en premier lieu affecter la durée et le type de scolarité suivi par ses bénéficiaires et provoque donc un supplément de dépenses pour la collectivité qu'il faut intégrer comme un bénéfice négatif. Les coûts d'une éducation secondaire (jusqu'au niveau de la terminale) sont identiques pour tous les participants, les données ne faisant pas apparaître de différence significative entre ceux ayant bénéficié du programme et les autres ; Heckman et al. (2010) ont évalué à 6 645\$ le coût d'une année d'enseignement secondaire dans l'état du Michigan durant les années concernées. Ils ont estimé ensuite les frais moyens d'inscription dans l'enseignement supérieur à l'aide de ceux observés en 1985 dans le Michigan ; cela donne 29\$ pour une heure de crédit²⁴⁷ au cours des deux premières années d'études supérieures, et 42\$ pour une heure de crédit au cours

²⁴⁶ Heckman J. J., Moon S. H., Pinto R., Savelyev P. A. et Yavitz A. (2010), "The rate of return to the High-Scope Perry Preschool Program", *Journal of Public Economics*, Vol. 94(1-2), p. 114-128.

²⁴⁷ Une heure de crédit est l'unité de mesure du crédit éducatif, généralement basée sur le nombre d'heures de cours par semaine pendant un trimestre.

des deux années suivantes. La dépense publique pour l'enseignement supérieur a été évaluée à 590\$ par heure de crédit au cours des deux premières années d'études supérieures, et 1 765\$ pour une heure de crédit au cours des deux années suivantes. Ces coûts ont servi à calculer les coûts relatifs au temps passé dans l'enseignement supérieur pour chaque individu (traité ou contrôle) puisque l'enquête de suivi enregistre son nombre d'heures de crédit.

Les bénéficiaires du programme

En l'absence d'observations au-delà de l'âge de 40 ans, Heckman *et al.* (2010) ont utilisé deux sources statistiques pour imputer les salaires perçus entre 40 et 65 ans : le *Current Population Survey* de mars 2002²⁴⁸ et le *Panel Study of Income Dynamics* (PSID). Ils ont estimé à l'aide de ces deux sources un modèle avec effets aléatoires, stratifié par genre et origine raciale, dans lequel les évolutions individuelles de salaires sont régressées sur le niveau d'éducation, l'âge et les salaires passés. Ce modèle leur a ensuite permis de prévoir pour chaque individu ayant participé au programme le salaire moyen qu'il ou elle aurait dû percevoir chaque année entre 40 et 65 ans. Les observations manquantes sur les salaires perçus avant 40 ans sont imputées à partir d'un modèle de régression estimé à l'aide du *National Longitudinal Survey of Youth* de 1979 (NLSY79). Heckman *et al.* (2010) ont appliqué à chaque salaire un taux d'imposition annuel de 15 % et un taux de cotisations sociales de 7,5 % (conformément au *Federal Insurance Contributions Act*). Ces impôts et cotisations, qui font partie des bénéfices sociaux du programme, réduisent les salaires nets perçus par les participants.

La contribution la plus importante d'Heckman *et al.* (2010) consiste en la monétisation des coûts de la délinquance que le programme a permis de réduire. Il n'est pas possible ici de détailler la procédure par laquelle ces coûts évités (qui sont inclus dans les bénéfices du programme) sont calculés. Les principaux ingrédients en seront toutefois rappelés. **L'appariement avec des sources judiciaires a permis de connaître avec précision le nombre d'arrestations, d'accusations, de condamnations, et d'incarcérations de chaque participant au programme.** Au-delà de l'âge de 40 ans, les nombres d'arrestations ont été extrapolés à partir des statistiques publiées par le *Federal Bureau of Investigation*. Heckman *et al.* (2010) ont calculé le coût social V_t^c d'un délit de type c commis à la date t comme le produit de trois termes :

$$V_t^c = C_t^c \times \left(\frac{I_t^c}{A_t^c} \right) \times A_t^c$$

Dans cette expression, le ratio I_t^c/A_t^c représente la proportion d'arrestations (*incidence-to-arrest ratio*) correspondant au délit de type c l'année t . Ce ratio est calculé à un niveau local (*i.e.*, les métropoles du Midwest) à partir de deux sources, le *Uniform Crime Report* (UCR) publié par le *Federal Bureau of Investigation* et le *National Crime Victimization Survey* (NCVS). Le terme A_t^c est observé directement : il correspond pour chaque participant au nombre d'arrestations pour des délits de type c au cours de l'année t .

²⁴⁸ Les individus ayant participé au programme Perry Preschool à l'âge de 5 ans en 1967 ont 40 ans en 2002.

Les coûts C_t^c d'un délit de type c ont deux composantes, les coûts supportés par la (ou les) victime(s) et les coûts supportés par le système judiciaire, ces derniers incluant les frais de police, de justice et d'incarcération. Les coûts supportés par les victimes sont décrits dans un ouvrage de Cohen (2005).²⁴⁹ Lorsque le délit est un meurtre (quatre cas recensés dans le groupe des participants, dont trois dans le groupe de contrôle), la valeur de la vie de la victime est estimée par Heckman *et al.* (2010) en utilisant alternativement deux scénarios, une limite basse (13 000\$) et une limite haute (4,1 millions de dollars). Les frais de police et de justice sont estimés à l'aide des informations fournies par le *Uniform Crime Report* (UCR) et par les fichiers de données individuelles *Expenditure and Employment Data for the Criminal Justice System* (CJEE). Les coûts d'incarcération sont calculés à partir d'informations fournies par les établissements pénitentiaires du Michigan.

La plupart des participants au programme *Perry Preschool* ont significativement bénéficié, à différents moments de leur vie, des divers programmes d'aide sociale. Pour calculer le montant de ces aides, Heckman *et al.* (2010) ont dû faire face à deux types de difficultés. Premièrement, ils ne disposaient pas d'informations suffisantes sur les aides en nature, celles concernant la santé, le logement, ou l'éducation des enfants. Par ailleurs, ils n'avaient que des informations très partielles sur les aides financières fournies par les programmes d'aide sociale *General Assistance* et *Aid to Families with Dependent Children* (devenu depuis 1996 *Temporary Assistance for Needy Families*), ou sur les montants perçus au titre de l'assurance-chômage. Pour les âges compris entre 27 et 40 ans, Heckman *et al.* (2010) ont utilisé les données du NLSY79 pour imputer les montants annuels d'aide sociale à ceux des participants qui y sont éligibles. Au-delà de 40 ans, les imputations sont faites à l'aide des fichiers du PSID. Les aides en nature sont simulées à l'aide des données du *Survey of Income and Program Participation* (SIPP).

Le taux de rendement interne et le ratio bénéfices/coûts

Le taux de rendement interne et le ratio bénéfices/coûts du programme *Perry Preschool* sont calculés pour chaque genre, puis pour l'ensemble des personnes, en fonction de plusieurs hypothèses. Comme indiqué précédemment, Heckman *et al.* (2010) ont utilisé deux chiffrages pour évaluer une vie humaine en cas de meurtre : la première (dite « valeur haute »), égale à 4,1 millions de dollars, est une valeur souvent utilisée dans la littérature académique, la seconde (dite « valeur basse ») correspond au coût induit par une agression violente, estimé à 13 000 dollars. Cette seconde hypothèse a été utilisée afin de ne pas surestimer les effets monétaires des meurtres imputables aux membres du groupe de contrôle dans l'évaluation finale, et donc les bénéfices associés à la participation au programme qui réduit la propension à commettre de tels actes (cf. Tableau 5).

La perte fiscale « sèche » (« *deadweight loss* ») est difficile à évaluer avec précision. C'est pour cette raison qu'Heckman et ses co-auteurs considèrent trois valeurs, 0,50 et 100 %, deux de ses valeurs (0 et 100 %) bornant l'espace des possibles. Leur interprétation finale et notamment leurs calculs des ratios bénéfices/coûts reposent sur la valeur médiane de 50% (cf. Tableau 5).

²⁴⁹ Cohen M. A. (2005), *The Costs of Crime and Justice*, Routledge, New York.

Les valeurs des taux de rendement interne pour un individu et pour la société, calculés sous ces différentes hypothèses, sont reportées dans le Tableau 5. Rappelons que le taux de rendement interne est défini comme le taux d'actualisation qui égalise le coût de mise en œuvre du programme et la différence entre les flux de bénéfices engendrés sur l'ensemble des périodes séparant l'âge d'entrée dans le programme de l'horizon retenu.²⁵⁰ Le taux de rendement interne ne peut être le seul critère d'efficacité considéré, il doit être complété par le ratio entre bénéfices et coûts actualisés qui est une mesure plus simple à obtenir. Dans le cas du programme *Perry Preschool*, les taux de rendement interne obtenus pour une valeur médiane de la perte fiscale « sèche » (50 %) sont de l'ordre de 7 % lorsqu'ils sont calculés pour une femme ou un homme pris individuellement, et de 8 % à 9 % pour la société prise dans son ensemble.²⁵¹ Ces valeurs sont proches du ratio bénéfices/coûts calculé pour l'ensemble de la société avec un taux d'actualisation égal à 3 %, ratio qui est égal en ce cas à 7,1. Quelles que soient les hypothèses faites sur l'importance de la perte fiscale et sur la valeur du taux d'actualisation, la mesure d'efficacité retenue (i. e., le taux de rendement interne ou le ratio bénéfices/coûts) est positive et statistiquement significative. Elle décroît légèrement lorsque la valeur de la perte fiscale « sèche » augmente, plus fortement lorsque le taux d'actualisation passe de 0 à 7 %.

Que retenir ?

- Évaluation coûts-bénéfices ex post d'un programme en faveur de la petite enfance sur le cycle de vie
- Basé sur une expérience aléatoire contrôlée et suivi longitudinal très long
- Apparié à des sources administratives pour les participants directs (e.g. sources judiciaires)
- Evaluation sérieuse des coûts évités (notamment les coûts de la délinquance)
- Appariements statistiques à d'autres sources pour reconstruire les trajectoires
- Prise en compte de l'incertitude des coefficients et calculs des erreurs-types associées
- Analyses de sensibilité très détaillées et riches

²⁵⁰ Ici cet horizon est égal à $T = 65 - 3 = 62$. Cela signifie que le taux de rendement interne est la solution d'un polynôme de degré 62, degré à l'évidence très élevé. Une telle équation peut bien évidemment avoir plusieurs solutions. Ce n'est pas le cas ici, Heckman *et al.* (2010, note de bas de page 73, p.124) affirmant que la solution qu'ils obtiennent est toujours unique et positive, quelle que soit l'hypothèse retenue.

²⁵¹ Lorsqu'ils calculent le rendement individuel, Heckman *et al.* (2010) tiennent seulement compte des gains recueillis et des coûts supportés directement par le seul bénéficiaire de l'intervention. Le rendement social est un indicateur plus large : il inclut l'impact du programme sur la délinquance, la participation à l'aide sociale, et les économies (ou suppléments) de coûts et transferts qui en résultent pour l'ensemble de la société.

**Tableau 5 – Taux de rendement interne et ratio bénéfices-coûts
du programme Perry Preschool**

Rendement	Pour un individu			Pour la société			Pour la société			
		Ensemble	Hommes	Femmes	Valeur haute (\$4,1M)	Hommes	Femmes	Valeur basse (\$13K)	Hommes	Femmes
Coût d'un meurtre										
Taux de rendement interne										
Perte fiscale	0 %	7,6	8,4	7,8	9,9	11,4	17,1	9,0	12,2	9,8
		(1,8)	(1,7)	(1,1)	(4,1)	(3,4)	(4,9)	(3,5)	(3,1)	(1,8)
	50 %	6,2	6,8	6,8	9,2	10,7	14,9	8,1	11,1	8,1
		(1,2)	(1,1)	(1,0)	(2,9)	(3,2)	(4,8)	(2,6)	(3,1)	(1,7)
	100 %	5,3	5,9	5,7	8,7	10,2	13,6	7,6	10,4	7,5
		(1,1)	(1,1)	(0,9)	(2,5)	(3,1)	(4,9)	(2,4)	(2,9)	(1,8)
Ratio bénéfices/coûts										
Taux d'actualisation	0 %				31,5	33,7	27,0	19,1	22,8	12,7
					(11,3)	(17,3)	(14,4)	(5,4)	(8,3)	(3,8)
	3 %				12,2	12,1	11,6	7,1	8,6	4,5
					(5,3)	(8,0)	(7,1)	(2,3)	(3,7)	(1,4)
	5 %				6,8	6,2	7,1	3,9	4,7	2,4
					(3,4)	(5,1)	(4,6)	(1,5)	(2,3)	(0,8)
	7 %				3,9	3,2	4,6	2,2	2,7	1,4
					(2,3)	(3,4)	(3,1)	(0,9)	(1,5)	(0,5)

Remarques : a) les écarts-types sont calculés par bootstrap ; ils sont reportés sous chaque estimation entre parenthèses ; b) dans le calcul du ratio bénéfices/coûts, la perte fiscale (« deadweight loss ») est supposée être égale à 50 % ; c) la valeur d'une vie humaine est supposée être égale à 4,1 millions de dollars (valeur haute, fréquemment utilisée), ou à 13 000 dollars (valeur très « basse », minimum extrême de ce qui peut être envisagé).

Source : Heckman et al. (2010), p.115

Annexes B

Quelques institutions produisant des revues systématiques de littérature

Plusieurs institutions ou organisations ont contribué à recenser de façon rigoureuse les évaluations d'impact de politiques ou d'interventions dans les champs de l'éducation, de l'emploi et de la santé. Nous présentons ici quelques sources pouvant être mobilisées pour mettre en œuvre des analyses coûts-bénéfices *ex ante* d'investissements sociaux.

Il ne s'agit pas d'un registre exhaustif et le fait que nous présentions certains centres ne discrédite en rien d'autres qui feraient un travail similaire.

Annexe B-I

Campbell Collaboration

<https://www.campbellcollaboration.org>

Campbell Collaboration est un réseau international de recherche fondé en 2000 et basé à Oslo. Son but est de promouvoir les politiques économiques et sociales fondées sur les résultats de la recherche en publiant des revues systématiques de littérature.

Les travaux publiés concernent **l'éducation, le handicap, la justice et la lutte contre la criminalité, le développement international et les politiques sociales**. Fin 2017, 144 revues systématiques ont été publiées²⁵².

Le processus de publication se déroule en trois temps. Tout d'abord, les auteurs enregistrent un « titre » auprès de Campbell Collaboration *via* un formulaire dédié. Puis ils développent un protocole de recherche – une sorte de plan de préanalyse – qui indique précisément les méthodes employées pour chaque étape de l'analyse, les sources mobilisées, les modèles statistiques utilisés dans le cadre de la méta-analyse, les critères d'inclusion et d'exclusion des travaux recensés, etc. Enfin, les auteurs soumettent la revue réalisée en suivant le protocole pré-enregistré. À chaque étape, une équipe de Campbell Collaboration spécialisée dans la thématique retenue adresse des remarques et des conseils à l'équipe de recherche. De plus, un processus de validation externe du protocole et de la revue est confié à des pairs : au moins deux chercheurs extérieurs sont sollicités pour expertiser le protocole de façon anonyme, l'un est choisi pour son expertise sur la thématique, l'autre pour son expertise méthodologique.

Des outils d'aide à la rédaction et à la constitution de la revue sont fournis par Campbell, ce qui permet de standardiser le contenu des revues publiées.

²⁵² Campbell Collaboration (2018), *Annual report 2017*, Oslo.

Annexe B-II

Washington State Institute for Public Policy

<http://www.wsipp.wa.gov/Publications>

Les travaux de ce centre ont été décrits en détail dans la section **Erreur ! Source du renvoi introuvable.**

Le *Washington State Institute for Public Policy* (WSIPP) produit des analyses coûts-bénéfices de politiques publiques qui pourraient être mises en place dans l'État de Washington. Pour y parvenir, il a développé un modèle d'analyse coûts-bénéfices complet à partir d'un ensemble de procédures claires et applicables à de nouvelles politiques. Ce modèle est alimenté par différents paramètres. Les coûts des interventions sont calculés comme si ces interventions devaient être mises en place dans l'État de Washington. Les effets des politiques testées sont calibrés à partir de méta-analyses. Celles-ci sont réalisées à partir de la littérature académique américaine, en prenant en compte des critères définis dans le document technique disponible sur le site.

Annexe B-III

Cochrane

<https://www.cochrane.org>

Cochrane est une organisation non-gouvernementale collaborative visant à produire des revues systématiques de littérature dans le champ de la santé. Fondé il y a vingt-cinq ans, le registre de revues systématiques de Cochrane atteint aujourd'hui 7 500 publications.

Cochrane a inspiré le développement de la Campbell Collaboration et les deux institutions collaborent largement dans la production de supports méthodologiques et dans l'organisation d'événements.

Cochrane vise à diffuser le plus largement possible les résultats des recherches. Pour cela, elle traduit en 14 langues les résumés des littératures des revues.

Annexe B-IV

Education Endowment Foundation

<https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit/>

L'*Education Endowment Foundation* fait partie du réseau des *What Works Centers* britanniques, centres de ressources visant à promouvoir et à développer des politiques et pratiques fondées sur la recherche et les données probantes. Cette organisation mène de nombreux travaux et finance notamment de nombreuses expérimentations aléatoires contrôlées en Grande Bretagne. Elle a par ailleurs développé une boîte à outils pour l'enseignement. Cette « *teaching learning toolkit* » est en réalité un registre de méta-

analyses de politiques éducatives et d'analyses coûts-efficacité, présentées de façon simple à partir de trois critères : le coût de l'intervention, l'impact et la qualité de la preuve.

Ce dernier critère est basé sur l'abondance des travaux utilisés pour réaliser la méta-analyse et sur la qualité de ces recherches. Contrairement aux autres institutions, l'*Education Endowment Foundation* est un peu moins exigeante lorsqu'il s'agit de sélectionner les travaux inclus dans les méta-analyses, mais elle pondère les études en fonction de leur méthodologie. Les impacts sont présentés de façon intuitive, par exemple sous la forme des nombres de mois gagnés, ce qui n'est en réalité qu'une façon de normaliser la taille d'un effet méta-analytique par la moyenne des progrès habituels des élèves au cours d'une année.

Annexe B-V

What Works Clearing House

<https://ies.ed.gov/ncee/wwc/>

Les *What Works Clearing House* sont en quelque sorte l'équivalent américain de l'*Education Endowment Foundation* dans le sens où ils recensent sous forme de revues systématiques et de méta-analyses les interventions et pratiques efficaces en éducation. Il s'agit d'un Département de l'*Institute for Education Science*, l'organisme statistique et de recherche du *US Department of Education*, le ministère de l'Éducation américain.

Annexe B-VI

National institute for health and care excellence (NICE)

<https://www.nice.org.uk/>

Le NICE a été créé en 1999 en Grande-Bretagne comme autorité indépendante visant à améliorer la qualité et la quantité de traitements et soins disponibles à partir de recherches de qualité. Il produit en particulier des guides et conseils basés sur des revues systématiques de littérature à destination des praticiens et décideurs en santé.

Annexes C

Quelques références de méta-analyses

Un certain nombre de revues systématiques de littérature de grande qualité ont été réalisés et permettent d'utiliser leurs résultats pour mener des analyses coûts-bénéfices. Nous recensons ici quelques travaux pour des politiques souvent évoquées dans le débat public dans les trois champs évoqués dans ce rapport. Ces revues sont principalement extraites des publications de *Campbell Collaboration*.

Annexe C-I

Quelques méta-analyses de l'impact de politiques d'emploi

Méta-analyse de l'effet de la réduction de la durée des droits à l'assurance chômage

Filges T., Jonassen A. B., Jørgensen A. M. K. (2018), "Reducing unemployment benefit duration to increase job finding rates: a systematic review", *Campbell Systematic Reviews* 2018:2

Cette méta-analyse est fondée sur les travaux publiés avant Mars 2016, dans lesquels un groupe de contrôle était bien défini. 41 études sur 34 000 répondaient aux exigences des critères d'inclusion, mais, parmi ces 41 études, seules 38 fournissaient suffisamment d'informations pour être incorporées dans la méta-analyse ; 28 d'entre elles présentaient un risque élevé de biais et 3 autres n'ont pu être intégrées car elles utilisaient le même échantillon. Finalement, la méta-analyse exploite 7 études. Les résultats indiquent qu'une réduction de la durée d'assurance chômage a un effet positif, significatif mais faible sur la probabilité de retrouver un emploi. Le taux de sortie vers l'emploi est augmenté d'en moyenne 10 % par rapport à ceux qui ont une durée de chômage plus longue. En d'autres termes, la probabilité qu'une personne dont la durée des droits est réduite trouve un emploi avant une personne disposant d'une durée de droits normale. Les données ne permettent pas de conclure sur la probabilité de réemploi ou sur le taux de salaire retrouvé.

Méta-analyse de l'effet d'Interventions pour améliorer l'activité des jeunes

Kluve J., Puerto S., Robalino D., Romero J. M., Rother F., Stöterau J., Weidenkaff F., Witte M. (2017), "Interventions to improve the labour market outcomes of youth: a systematic review of training, entrepreneurship promotion, employment services, and subsidized employment interventions", *Campbell Systematic Reviews* 2017:12.

Cette méta-analyse est fondée sur les travaux publiés avant janvier 2015 ; elle concerne les politiques actives sur le marché du travail dont les bénéficiaires sont âgés de 15 à 35 ans. Ces politiques peuvent être des politiques de **formation ou de développement de compétences**, des programmes **d'accompagnement à l'entrepreneuriat**, des **aides à la recherche d'emploi** ou **des emplois aidés**. Les travaux retenus reposent soit sur une expérience aléatoire contrôlée soit sur des évaluations quasi-expérimentales. Les résultats devaient porter sur l'emploi et le revenu des individus ou sur la rentabilité des entreprises.

L'analyse a sélectionné 113 études sur 32 000 disponibles. Ces études concernent 31 pays ; 55 concernent des formations, 15 des accompagnements à l'entrepreneuriat, 10 des aides à la recherche d'emploi, le restant porte sur les emplois aidés ; 47 % des études sont fondées sur des expériences aléatoires contrôlées. De façon générale, les effets sur l'emploi et les revenus des bénéficiaires sont positifs, mais il existe une forte hétérogénéité des effets des programmes. La promotion de l'entrepreneuriat et la formation ont des effets positifs, mais les services à la recherche d'emploi et les emplois aidés n'ont pas d'effet significatif.

Méta-analyse des politiques actives d'emploi en faveur des bénéficiaires de l'assurance chômage

Filges T., Smedslund G., Knudsen A. S. D., Jørgensen A. M. K. (2015), "Active Labour Market Programme Participation for Unemployment Insurance Recipients: A Systematic Review", *Campbell Systematic Reviews* 2015:2.

Cette revue retient 73 études parmi 16 000 en respectant les critères d'inclusion habituels (notamment, disposer d'un groupe de contrôle bien défini). Elles proviennent de 15 pays. Parmi elles, seules 47 fournissent les éléments nécessaires aux calculs méta-analytiques mais six d'entre elles ont été exclues à cause du risque de biais. Finalement, la méta-analyse repose sur 39 études. Les résultats indiquent un effet général positif mais variable en fonction des méthodes d'analyses. L'impact n'apparaît pas significatif s'il est mesuré à une date donnée, et il apparaît positif si l'on tient compte des estimations des taux instantanés de sortie. L'effet global de la participation à une politique active correspond à une différence de probabilité de 7 points, ce qui signifie qu'en moyenne, il faut accompagner 15 personnes pour qu'une soit en emploi 1 an plus tard.

Ces travaux sont globalement en accord avec la méta-analyse plus récente de Card *et al.* (2018) détaillée dans la section (3.II.2c) :

D. Card, J. Kluve et A. Weber (2018), "What works? A meta-analysis of recent active labor market program evaluations", *Journal of the European Economic Association*, Vol 16(3), p. 894-931

Annexe C-II

Quelques méta-analyses de politiques éducatives

Méta-analyse de politiques de prévention du décrochage scolaire

Wilson S. J., Tanner-Smith E. E., Lipsey, M. W., Steinka-Fry, K., Morrison, J. (2011), "Dropout prevention and intervention programs: Effects on school completion and dropout among school aged children and youth", *Campbell Systematic Reviews* 2011:8.

Cette revue analyse les recherches publiées entre 1985 et 2010 dont le design est soit expérimental soit quasi expérimental. 167 études sur 24 000 ont été retenues et séparées en deux groupes : 152 études sont incluses dans une méta-analyse de programmes

généraux de lutte contre le décrochage, les 15 études restantes concernent des programmes à destination de parents adolescents (*teen parents*).

Les résultats de la première méta-analyse indiquent une baisse du risque de décrochage de 72 % (soit une proportion de décrocheurs passant de 21,1 % à 13 %). Pour les programmes à destination des parents adolescents, la baisse du risque de décrochage est de 83 %.

Méta-analyse de l'effet de programmes relatifs aux activités extra-scolaires

Zief S. G., Lauver S., Maynard R. A. (2006), "Impacts of after-school programs on student outcomes", *Campbell Systematic Reviews* 2006:3.

Cette revue est uniquement fondée sur des travaux expérimentaux de programmes qui intègrent à la fois de l'accompagnement éducatif et d'autres activités plus récréatives. Seules cinq études ont pu être retenues ; elles concernent des programmes très similaires. L'effet de ces programmes sur la probabilité de participer à une activité artistique ou sportive est positif et significatif (10 % d'un écart type). Mais l'effet sur les compétences scolaires n'apparaît pas comme significatif.

Méta-analyse de l'effet d'une baisse de la taille des classes au primaire

Filges T., Sonne-Schmidt C. S., Nielsen B. C. V. (2018), "Small class sizes for improving student achievement in primary and secondary schools", *Campbell Systematic Reviews* 2018:10.

Cette étude a été présentée en détail dans un encadré de la section 2.IV.3.

Recension et synthèse de 196 expérimentations aléatoires contrôlées en éducation

Roland G. Fryer, Jr (2016). "The Production of Human Capital in Developed Countries: Evidence from 196 Randomized Field Experiments », *NBER Working Papers* 22130, National Bureau of Economic Research, Inc.

Cet article recense de façon assez exhaustive les résultats de recherches basées sur des expérimentations aléatoires contrôlées menées dans des pays développés et pour lesquelles il est possible de vérifier qu'une affectation aléatoire a bien été mise en œuvre, en dehors de contextes de laboratoires. L'auteur analyse 196 études qu'il divise en trois catégories : les interventions précoces, les interventions à l'école et les interventions au domicile familial. L'auteur réalise ensuite des méta-analyses pour ces trois groupes d'interventions (assez peu comparables) et indique un effet moyen de :

- 11,1 % d'un écart type en mathématiques et 16,5 % d'un écart-type en lecture pour les politiques relatives à la petite enfance ;
- 5,2 % et 6,8 % respectivement pour les interventions à l'école ;
- Pas d'effets significatifs pour les interventions au domicile familial.

Annexe C-II

Quelques revues systématiques de politiques de prévention en santé

Comment inciter les enfants à manger davantage de fruits et légumes

Knai C., Pomerleau J., Lock K., McKee M., (2006), "Getting children to eat more fruit and vegetables: a systematic review", *Prev Med.*; Vol. 42(2), p. 85-95.

Cette revue recense les travaux portant sur des politiques visant à inciter les enfants bien portants à consommer plus de fruits et de légumes. Elle retient les études réalisant un suivi d'au moins trois mois, et incluant un groupe de contrôle. Les variables de résultats se focalisent sur la consommation de fruits et légumes tout en faisant référence à la littérature indiquant que leur consommation est associée à une réduction des problèmes de santé.

Quinze travaux sur 3 500 respectaient les critères d'inclusion. Aucun n'a trouvé un impact négatif sur la consommation de légumes et de fruits, dix ont conclu à des effets positifs, la consommation moyenne ayant augmenté de 0,3 à 0,9 fruits et légumes en plus par jour.

L'efficacité et l'efficience de politiques visant à réduire les dommages causés par l'alcool

Anderson P., Chisholm D., Fuhr D. C. (2009), "Effectiveness and cost-effectiveness of policies and programmes to reduce the harm caused by alcohol", *Lancet*, Vol.373(9682), p.2234-46.

Ce travail propose une synthèse des résultats des politiques visant à réduire les dommages causés par l'alcool. Il est réalisé à partir des revues systématiques publiées par *Cochrane*. Ces dommages peuvent être de natures diverses : blessures et violences (intentionnelles ou non) à soi ou à autrui, réduction des performances au travail, absentéisme, suicide, meurtre, crime, accidents de la route, etc.

Le niveau de preuve des diverses politiques testé est classé en six catégories : 1) s'il existe plusieurs revues systématiques, 2) s'il existe une revue systématique, 3) s'il existe plusieurs expériences aléatoires contrôlées, 4) s'il existe une évaluation aléatoire, 5) s'il existe des études observationnelles et 6) s'il n'y a pas d'évaluation.

Les politiques concernent plusieurs types d'intervention :

- Information et éducation
- Politiques de santé publique
- Programmes communautaires
- Programmes de prévention contre l'alcool au volant
- Limite de l'accessibilité de l'alcool
- Contraintes sur le marketing des produits alcoolisés
- Politiques de prix
- Réduction des dommages

Les politiques qui apparaissent particulièrement efficaces sont celles régulant l'environnement dans lequel l'alcool est commercialisé, que ce soit par une régulation du prix, du marketing ou de la réglementation de la vente.

Après prise en compte des coûts des politiques évaluées, il apparaît que les plus efficaces sont celles qui consistent à augmenter le prix de vente ainsi que la mise en place de réglementations limitant la publicité. Les politiques de prévention en milieu scolaire n'ont pas d'effet sur la réduction des dommages, mais les programmes d'information et d'éducation peuvent augmenter l'attention du public sur ces enjeux.

L'effet d'une combinaison d'activité physique et d'une alimentation adaptée sur la prévention du diabète de type II

Li R., Qu S., Zhang P., Chattopadhyay S., Gregg E. W., Albright A., *et al.* (2015), "Economic Evaluation of Combined Diet and Physical Activity Promotion Programs to Prevent Type 2 Diabetes Among Persons at Increased Risk: A Systematic Review for the Community Preventive Services Task Force", *Ann Intern Med.*, Vol. 163(6), p. 452-60.

Ce travail présente une revue systématique des coûts, impacts et bénéfices de programmes d'activité physique et/ou d'un régime diététique adapté visant à limiter le risque de diabète de type 2 parmi des populations à risque²⁵³. 28 études ont été prises en compte. Le coût moyen de ces programmes est estimé à 653 \$₂₀₁₃ ; 22 études ont estimé le ratio coûts-efficacité de ces programmes ; 16 rapportent une mesure par QALY gagné. Le ratio médian est ainsi de 13 761\$₂₀₁₃/QALY gagné. Aucune évaluation coûts- bénéfices n'a été recensée.

²⁵³ C'est-à-dire un diabète non génétique.

Annexe D

Liste des membres du groupe de travail

Président

Denis Fougère, directeur de recherches, CNRS, Sciences Po

Coordinateur

Arthur Heim, chef de projets, France Stratégie

Intervenants

Pierre Arwidson, directeur adjoint de la prévention et de la promotion de la santé, Santé Publique France

Luc Baumstark, Secrétariat général pour l'investissement

Clément Belet, London School of Economics

Didier Blanchet, Insee

Marine Boisson-Cohen, conseillère synthèses et prospective, cabinet du Haut-Commissaire à la réforme des retraites

Pierre Cahuc, Crest

Gabrielle Fack, Université Paris Dauphine

Marc Ferracci, conseiller spécial, ministère du Travail

Roberto Galbiati, directeur de recherche, CNRS-Sciences Po

Cecilia Garcia-Penalosa, université d'Aix Marseille

Robert Gary-Bobo, Sciences Po Paris

Pauline Givord, OCDE

Marc Gurgand, Paris School of Economics

Nicolas Jacquemet, Paris School of Economics

Florence Jusot, Université Paris Dauphine

Fabrice Lengart, commissaire général adjoint, France Stratégie

Christine Leclainch, université de Lille II

Jincheng Ni, chef de projets, France Stratégie

Cyril Nouveau, directeur des statistiques, des études et de l'évaluation, Pôle Emploi

Roland Rathelot, University of Warwick

Lise Rochaix, Paris School of Economics

Anne-Laure Samson, professeur, université de Lille

Claudia Senik, Paris school of Economics

Clémence Thébaut, université Paris Dauphine

Nicolas Treich, Toulouse School of Economics

Marie-Claire Villeval, CNRS Lyon

Jérôme Wittwer, université de Bordeaux

Membres du groupe

Morad Ben Mezian, chargé de missions, France Stratégie

Emmanuel Bretin, DG Trésor

Pauline Charnoz, Dares

Rozenn Desplatz, France Stratégie

Morgane Gaudiau, Dares

Claude Gissot, CNAM

Sylvain Grognet, Dares

Eric Kulanthaivelu, Dares

Mahaut Lasselin, Dares

Gautier Maigne, chef de département, France Stratégie

Carine Milcent, Paris School of Economics

Olivier Monso, Depp, ministère de l'Éducation nationale et de la Jeunesse

Antoine Naboulet, France Stratégie

Romain Roussel, adjoint à la sous-directrice de DRESS

Julien Rousselon, adjoint au chef de département, France Stratégie

Annexe E

Déroulé des séances du groupe de travail

Séance inaugurale, 2 février 2018

- **Propos introductifs** – Fabrice Lenglard (France Stratégie)
- **Problématisation et organisation** – Denis Fougère, président du groupe (CNRS, Sciences Po)
- **L'investissement social : quelle stratégie pour la France ?** – Marine Boisson-Cohen (Haut-Commissariat à la réforme des retraites)
- **Le guide de l'évaluation socioéconomique des investissements publics** – Jincheng Ni (France Stratégie)
- **Peut-on estimer le rendement de l'investissement social** – Arthur Heim, coordinateur du groupe (France Stratégie)
- **Avantages et limites des évaluations d'impact** – Denis Fougère, président du groupe (CNRS, Sciences Po)

Enjeux méthodologiques et considérations pratiques, 5 Mars 2018

- **La mise en œuvre du calcul socioéconomique en santé** – Lise Rochaix (Paris School of Economics) et Jérôme Wittwer (Université de Bordeaux)
- **Le calcul socioéconomique au défi de la mesure de l'utilité sociale des investissements publics** – Luc Baumstark (Secrétariat général pour l'investissement)
- **Quelques grandes difficultés (et petites subtilités) dans l'analyse coûts-bénéfices** – Nicolas Treich (Toulouse School of Economics)
- **Modèles de microsimulation : quelle complémentarité avec les autres outils d'évaluation des politiques sociales ?** – Didier Blanchet (Insee)
- **Les politiques d'aide à la recherche d'emploi** – Roland Rathelot (University of Warwick)

Investissement social et éducation, 4 avril 2018

- **La taille des classes influence-t-elle la réussite scolaire ?** – Marc Gurgand (Paris School of Economics)
- **Ségrégation scolaire : quel bilan possible des politiques d'affectation scolaire ?** – Pauline Givord (INSEE SSP-Lab)
- **Dans quelle mesure les contraintes financières limitent-elles l'accès à l'enseignement supérieur ? Le rôle des politiques d'aide aux étudiants** – Gabrielle Fack (Université Paris Dauphine)
- **Le financement de l'enseignement supérieur** – Robert Gary-Bobo (Crest)

Investissement social et emploi, 15 Mai 2018

- **The effectiveness of hiring credits** — Pierre Cahuc (Crest)
- **Chômage et délinquance, résultats et recommandations** — Roberto Galbiati (Cnrs, Sciences Po)
- **Evaluation socioéconomique de l'investissement social : Quelques éléments de réflexion à partir des politiques d'emploi** — Cyril Nouveau (Pôle Emploi)
- **La formation tout au long de la vie** — Marc Ferracci (Cabinet de la ministre du Travail)

Investissement social et santé, 6 juin 2018

- **Le recours à une valeur de référence du QALY dans le calcul économique en santé** – Clémence Thébaut (Université de Limoges)
- **Intégrer le revenu équivalent santé dans l'approche coûts-bénéfices traditionnelle : L'exemple des traitements antihypertenseurs** – Anne-Laure Samson (Université Paris Dauphine)
- **Les critères d'équité dans l'évaluation économique des politiques de santé** – Florence Jusot (Université Paris Dauphine)
- **La responsabilité individuelle et l'évaluation économique des politiques de santé** – Christine Le Clainche (Université de Lille 2)
- **Travaux sur l'efficacité de différentes interventions de prévention : exemples de systèmes d'aide à la décision** – Pierre Arwidson (Santé Publique France)

Investissement social et bien-être, 4 juillet 2018)

- **Politique d'égalisation hommes-femmes sur le marché du travail** – Cecilia Garcia-Penalosa (Aix-Marseille Université, AMSE)
- **UK Wellbeing Simulator : Exemple de microsimulation d'intervention au Royaume-Uni et résultats** – Clément Bellet (London School of Economics) et Claudia Senik (Paris School of Economics)
- **Les apports de l'économie comportementale et expérimentale à l'évaluation des politiques publiques** – Marie Claire Villeval (CNRS, Lyon)
- **Les apports de l'économie expérimentale à l'évaluation des politiques publiques : comparaison avec l'évaluation randomisée** – Nicolas Jacquemet (Paris School of Economics)

Restitution du rapport, 6 décembre 2018

- **L'évaluation socioéconomique de l'investissement social : comment mettre en œuvre des analyses coûts-bénéfices pour les politiques sociales ou éducatives ?** – Denis Fougère (Cnrs, Sciences Po Paris) et Arthur Heim (France Stratégie)



Directeur de la publication

Gilles de Margerie, commissaire général

Directeur de la rédaction

Cédric Audenis, commissaire général adjoint

Secrétaires de rédaction

Olivier de Broca, Valérie Senné

Contact presse

Matthias Le Fur, directeur du service Édition/Communication/Événements

01 42 75 61 37, matthias.lefur@strategie.gouv.fr

RETROUVEZ
LES DERNIÈRES ACTUALITÉS
DE FRANCE STRATÉGIE SUR :



www.strategie.gouv.fr



[@Strategie_Gouv](https://twitter.com/Strategie_Gouv)



[france-strategie](https://www.linkedin.com/company/france-strategie)



[FranceStrategie](https://www.facebook.com/FranceStrategie)



[@FranceStrategie_](https://www.instagram.com/FranceStrategie_)



[StrategieGouv](https://www.youtube.com/StrategieGouv)



FRANCE STRATÉGIE

Institution autonome placée auprès du Premier ministre, France Stratégie contribue à l'action publique par ses analyses et ses propositions. Elle anime le débat public et éclaire les choix collectifs sur les enjeux sociaux, économiques et environnementaux. Elle produit également des évaluations de politiques publiques à la demande du gouvernement. Les résultats de ses travaux s'adressent aux pouvoirs publics, à la société civile et aux citoyens.