

ANALYSES DU BIAIS DE NON-RÉPONSE AU NATIONAL CENTER FOR EDUCATION STATISTICS

Jonaki Bose¹

RÉSUMÉ

Dans les enquêtes dont le taux de réponse est bas, le biais de non-réponse peut constituer un grand sujet d'inquiétude. Il n'est pas toujours possible de mesurer le biais effectif, mais diverses méthodes aident à constater les sources possibles de cette déformation. Au National Center for Education Statistics (NCES), on procède à une analyse de biais de non-réponse pour toute enquête dont le taux de réponse est de moins de 70 %. Dans le présent document, il sera question des différents modes d'analyse de biais de non-réponse à l'aide d'exemples venant du NCES.

MOTS CLÉS : Non-réponse; Biais; Taux de réponse

1. INTRODUCTION

On constate habituellement que ce ne sont pas toutes les unités échantillonnées qui répondent à une enquête. Les probabilités de non-réponse augmentent s'il existe plusieurs stades ou volets de réponse p. ex. : interviews de présélection, répondants multiples dans un dossier ou cycles multiples de collecte de données. Ainsi, à l'enquête nationale sur l'éducation auprès des ménages (National Household Education Study ou NHES) du National Center for Education Statistics (NCES), qui est une enquête à composition aléatoire (CA), on effectue des interviews de présélection dont le taux d'achèvement est inférieur à celui des autres volets (Nolin et coll., 2000). De même, on constate qu'il a été plus difficile de persuader les établissements scolaires échantillonnés de participer au premier stade de l'enquête longitudinale sur la petite enfance pour les classes de maternelle en 1998-1999 (Early Childhood Longitudinal Survey: Kindergarten Class ou ECLS-K) et d'en obtenir la collaboration que de mobiliser les enquêtés visés à tout autre volet de cette étude (Brick, Burke et Lê, 2000). Cette absence de réponse d'unités échantillonnées peut contribuer à un biais dans les estimations de l'enquête.

Suivant les normes du NCES, si le taux général de réponse à une enquête (qui est le produit des taux d'achèvement de ses divers stades ou degrés) est de moins de 70 %, on doit effectuer une analyse de biais de non-réponse pour constater les sources de déformation des estimations par une non-réponse importante.

Le présent exposé, qui s'inspire des expériences des enquêtes du NCES auprès des ménages et des établissements primaires-secondaires et postsecondaires, examine la démarche de constatation des besoins et des caractéristiques d'une enquête qui influent sur la nature des analyses ensuite menées du biais de non-réponse. On y évalue aussi les buts, les forces et les faiblesses des différentes techniques employées. Notre propos n'est cependant pas de mettre l'accent sur les résultats de fond de telles analyses.

¹ Jonaki Bose, National Center for Education Statistics, 1990, rue K n.-o., Washington, D.C. 20006, États-Unis.

2. BIAIS DE NON-RÉPONSE

2.1 Définition et nécessité des analyses de biais de non-réponse

Le biais est la différence entre une estimation d'enquête et la valeur réelle de population. Dans une enquête par sondage, on peut y voir la valeur espérée de cette différence basée sur tous les échantillons possibles. Le biais de non-réponse lié à une estimation comprend deux éléments, à savoir l'ordre de grandeur de cette non-réponse et l'écart d'estimation entre répondants et non-répondants.

Le biais d'une estimation peut s'exprimer mathématiquement pour que soit dégagé le rapport entre ce biais et les deux facteurs précités. Le biais s'obtient par :

$$\text{Biais}(\hat{y}_r) = p_n \{E(\hat{y}_r - \hat{y}_n)\}$$

où \hat{y}_r est la caractéristique estimée à partir des répondants seulement, p_n est le taux de non-réponse, \hat{y}_n la caractéristique estimée à partir des non-répondants seulement et E l'espérance mathématique basée sur tous les échantillons possibles (Nolin et coll., 2000). Il peut y avoir biais par non-réponse d'unités ou non-réponse à des questions.

Ce biais a à voir tant avec les faibles des taux de réponse qu'avec l'importance des différences d'estimations entre répondants et non-répondants. Les estimations d'une étude risquent d'être entachées d'un biais par non-réponse à un ou plusieurs stades de l'enquête. La meilleure façon de prévenir une telle déformation est de relever les taux de réponse par des techniques de conversion intensive des cas de refus, des moyens d'incitation à la réponse, une pluralité de modes de collecte de données, une souplesse d'ordonnement et une formation des intervieweurs. Malgré tous les efforts, il peut y avoir non-réponse et, en pareil cas, les enquêtes ajustent les poids afin de compenser pour la non-réponse. Quelles que soient nos corrections de pondération en fonction de la non-réponse, les estimations peuvent toujours accuser un biais.

Il n'est pas toujours possible d'évaluer ce biais, car on ignore la valeur réelle du paramètre de population. Si on la connaît, la différence entre la valeur calculée à l'aide des données d'enquête et la valeur réelle pour la population peut être considérée comme une estimation du biais dont est entachée une estimation.

L'analyse du biais de non-réponse est l'exercice par lequel on estime le biais de non-réponse et en constate les causes. Elle permet une évaluation des statistiques d'une enquête estimées en utilisant les poids de base (calculé à partir des probabilités de sélection) et les poids ajustés pour la non-réponse.

Ces analyses sont utiles à divers égards. Elles servent de baromètre de la qualité des données recueillies et aident à relever les estimations susceptibles d'accuser un biais. Elles peuvent aussi aider à rassurer les utilisateurs des données, tout comme l'organisme qui recueille et diffuse cette information, au sujet de la qualité des données disponibles. En même temps, elles avertissent l'utilisateur de l'exposition de cette même information à des déformations. Grâce à elles, on peut mieux évaluer les variables utilisées dans les ajustements de pondération en fonction de la non-réponse. On peut également constater les sources de biais éventuels que l'on pourra éliminer dans de futurs cycles de collecte des données d'une étude longitudinale, laquelle sera particulièrement exposée au biais de non-réponse, le biais d'un premier cycle de collecte pouvant se retrouver dans de futurs cycles. Les enquêtes transversales répétées profitent aussi de telles analyses. Ainsi, on a soumis à une analyse de non-réponse les données de l'enquête de 1993-1994 sur les établissements et le personnel scolaires (Schools and Staffing Survey ou SASS; Monaco et coll., 1997). Cet exercice a donné lieu à de nombreuses recommandations pour les futures reprises de cette enquête. On en a tenu compte à la reprise de 1999-2000.

2.2 Facteurs influant sur les modes d'analyse de biais de non-réponse

Une diversité de facteurs déterminent le choix de modes d'analyse de biais de non-réponse. Il est bon de constater ces caractéristiques avant d'entreprendre une telle analyse.

Pluralité de volets : Sous sa forme la plus simple, une enquête comporte un seul questionnaire pour tous les enquêtés; on n'y trouve ni interviews de présélection (filtrage) ni volets multiples. Dans un certain nombre d'enquêtes cependant, on effectue des interviews de présélection qui peuvent déterminer l'admissibilité à une autre enquête. Ainsi, dans l'enquête NHES, on peut demander aux membres des ménages, d'après leurs réponses à un questionnaire de présélection, de répondre à une autre enquête portant, par exemple, sur l'éducation des adultes ou la participation à des programmes de la petite enfance ou à des programmes préscolaires ou postsecondaires. Dans d'autres enquêtes, chaque cas peut comporter plus d'un volet. Ainsi, dans l'enquête longitudinale sur la petite enfance pour les classes de maternelle de 1998-1999 (ECLS-K), l'enquêté principal est l'enfant qui subit une évaluation, mais on recueille aussi des données auprès des enseignants, des parents et des administrateurs des écoles. Il est important de tenir compte de ces différents volets et d'établir qui était admissible à compléter ces volets avant de procéder à une analyse du biais de non-réponse.

Caractère longitudinal ou transversal des enquêtes : Les modes d'analyse du biais peuvent être semblables pour le premier cycle d'une enquête longitudinale ou transversale. Par contre, l'évaluation du biais aux cycles ultérieurs de collecte d'une enquête longitudinale se doit de tenir compte du premier cycle.

Existence de poids multiples : Cela a à voir avec les deux premiers points, c'est-à-dire avec la pluralité de volets et la nature longitudinale des enquêtes. Nombre d'enquêtes ont plus d'un poids, même s'il s'agit d'études transversales. Préalablement à toute étude, il est nécessaire de déterminer quels poids il convient d'utiliser dans divers modes d'analyse, s'il est utile ou non d'évaluer les estimations basées sur plus d'un ensemble de poids, et quelles populations sont couvertes selon les poids utilisés.

3. MÉTHODES D'ANALYSE

3.1 Examen des taux de réponse

Comme nous l'avons précisé, le biais de non-réponse comprend deux éléments, à savoir l'ordre de grandeur de la non-réponse et la différence entre les résultats observés des répondants et les résultats inobservés des non-répondants. En règle générale, la première mesure qui peut aider à juger de la nécessité d'une évaluation plus poussée est l'examen du degré de non-réponse à une enquête.

Dans une enquête à un degré, il y a généralement un seul ensemble de taux de réponse d'intérêt pour l'étude de l'ordre de grandeur de la non-réponse. Dans une enquête à degrés ou à volets multiples, on peut utilement se reporter à plusieurs types de taux de réponse. Dans certaines enquêtes du NCES, les termes « taux d'achèvement » et « taux de réponse » désignent des notions différentes (Brick, Burke et Lê, 2000).

Les taux d'achèvement visent un pourcentage d'unités participantes à chaque degré de l'échantillonnage et se calculent séparément pour des volets et des questionnaires différents. Par taux de réponse, on entend le taux général de participation à une étude tous degrés d'échantillonnage confondus. Ainsi, dans l'enquête ECLS-K, le taux de réponse est le produit des taux de réponse des établissements scolaires (pourcentage des écoles qui ont consenti à participer à l'enquête) et du taux d'achèvement d'un volet. Ainsi, le taux de réponse pour le volet de l'évaluation des enfants sera le produit du taux de réponse des écoles et du taux d'achèvement de ce volet (pourcentage des enfants évalués étant donné la participation de leur école à l'étude). Les taux d'achèvement aident à relever les différences à l'intérieur de sous-groupes au même niveau, alors que les taux de réponse brossent le tableau d'ensemble, mais en risquant de brouiller le tableau des sources de biais.

Dans l'enquête SASS de 1993-1994, on a soumis les taux de réponse à une analyse très fine (Monaco et coll., 1997). Cette enquête étant à volets multiples (conseils scolaires locaux (« local education agencies » ou LEA), établissements scolaires, enseignants et élèves), on s'est attaché à la nature correspondante de la non-réponse. On s'intéressait dans ce cas aux « effets conjugués de la non-réponse ». On a voulu vérifier si les taux de réponse étaient indépendants selon qu'il s'agissait d'enseignants, de LEA ou d'administrateurs, d'établissements, de bibliothèques et de bibliothécaires scolaires du secteur public ou

du secteur privé. Ainsi, l'analyse a permis de constater que les enseignants des écoles privées présentaient un taux de réponse bien supérieur lorsque l'administrateur de leur école répondait à l'enquête.

Dans la plupart des enquêtes du NCES, on calcule les taux de réponse et d'achèvement sans les poids de même qu'avec les poids de base. aussi bien hors pondération que dans une pondération en fonction des seules probabilités de sélection (pondération de base). En général, l'évaluation de ces taux se fait avec les poids de base i.e. n'incluant aucun facteur d'ajustement.

L'évaluation des taux de réponse est le point de départ. Si on observe des taux de réponse élevés non seulement pour l'échantillon d'ensemble mais aussi pour des sous-échantillons, c'est peut-être l'indice qu'il est inutile de pousser l'analyse du biais de non-réponse. Comme nous l'avons signalé, tout taux général de réponse de moins de 70 % au NCES appelle une analyse de biais de non-réponse.

La plupart des taux de réponse qui se calculent visent des enquêtes, des degrés ou des volets entiers, mais la non-réponse liée à une estimation d'enquête est double, puisqu'il peut y avoir non-réponse d'unités et non-réponse à des questions. Dans certaines enquêtes du NCES, on étudie à la fois la non-réponse des unités de l'échantillon et la nature de la non-réponse à des questions.

Dans l'enquête longitudinale nationale sur l'éducation de 1988 (National Education Longitudinal Study ou NELLS), on a soumis la non-réponse à des questions à une vaste analyse (Spencer et coll., 1990). On a non seulement regardé les taux de réponse aux questions, mais aussi évalué dans ce cas des facteurs comme la disposition des questions sur le questionnaire, les sujets et la présence de filtres de cheminement de question en question. On a ensuite considéré les questions aux taux de non-réponse les plus élevés selon des caractéristiques des élèves comme le sexe, la race-ethnicité, la situation socio-économique et les résultats aux tests cognitifs. On s'est en outre attaché au nombre moyen de questions sans tentative de réponse aux tests cognitifs selon ces mêmes caractéristiques des élèves. Il existe d'autres méthodes d'évaluation de la non-réponse à des questions. Dans l'enquête longitudinale sur les études universitaires par exemple, on a considéré le lien entre les taux de réponse aux questions et les dates d'interview (Green et coll., 1999).

3.2 Comparaison des estimations d'échantillons de réponse avec les valeurs de population

En théorie, la meilleure façon de cerner le biais d'estimation à partir d'un échantillon de répondants est de comparer les estimations aux valeurs réelles de la population. Le plus souvent, on ne dispose pas des valeurs de la population. Cependant, certaines sources telles que les bases de sondage ou les dossiers administratifs peuvent fournir les valeurs de l'ensemble de la population ou une partie de celle-ci.

Dans les enquêtes auprès des établissements scolaires du NCES, le « Common Core of Data » (CCD) et l'enquête auprès des écoles privées (« Private School Survey » ou PSS) servent surtout de base de sondage pour les écoles publiques et privées respectivement. Ce sont là des enquêtes « universelles » avec des variables comme les effectifs scolaires totaux, les niveaux de scolarité et la répartition en pourcentage des minorités raciales ou ethniques dans la population enfantine scolarisée. Dans les enquêtes du NCES auprès des établissements postsecondaires, on s'est reporté à l'« Integrated Post-Secondary Educational System » (IPEDS) comme base de sondage. On y trouve des variables comme les effectifs, la régie (publique ou privée) et le plus haut niveau de fréquentation.

On peut comparer les estimations obtenues à partir des répondants des enquêtes-établissements (au double niveau des élèves et des écoles) aux valeurs de population pour dégager les biais. Un intervalle de confiance de la différence des estimations qui comprend la valeur zéro indique l'absence de biais.

Dans les enquêtes auprès des ménages, on n'a souvent guère de renseignements, surtout s'il s'agit d'une enquête à composition aléatoire. Dans ce cas, l'information disponible se limite surtout aux caractéristiques de circonscriptions téléphoniques et aux autres grandes caractéristiques géographiques pour chaque numéro de téléphone échantillonné. D'autres bases de sondage sont mises au service des enquêtes du NCES. Dans

le cas de l'enquête longitudinale sur la petite enfance par cohortes de naissances, les certificats ou actes de naissance servent surtout de base de sondage. Celle-ci est particulièrement riche en comparaison à d'autres bases, puisqu'elle livre non seulement des indications démographiques de base sur l'enfant et ses parents comme l'âge, le sexe et la race, mais renseigne aussi en détail sur la santé maternelle et les pratiques prénatales.

On peut établir des estimations pondérées en utilisant soit les poids de base soit le poids ajustés pour la non-réponse. Les estimations qui utilisent les poids non ajustés (poids de base) aident à évaluer le biais avant correction de non-réponse. Il y a des statisticiens qui préfèrent utiliser les poids de base, les données de la base de sondage même servant aux corrections de non-réponse. Il reste que les poids ajustés pour la non-réponse permettent une comparaison entre des estimations « définitives » et des valeurs de population.

Si l'examen de la différence entre estimations et valeurs de population nous donne un indicateur de biais, on ne se trouve pas à distinguer le biais d'échantillonnage du biais de non-réponse. On pourrait isoler le biais de non-réponse du biais d'ensemble, en évaluant le biais dont auraient été entachées les estimations si toutes les unités de l'échantillon avaient répondu. Ce serait le biais dû à l'échantillonnage. Une autre possibilité afin d'obtenir un indicateur direct du biais de la non-réponse est de comparer les estimations obtenues des répondants aux estimations des non-répondants.

3.3 Comparaison des estimations d'une enquête avec des estimations extérieures

C'est là une des voies où s'engagent le plus fréquemment tant les statisticiens que les chercheurs pour juger de la qualité des estimations d'une enquête. Il s'agit dans ce cas de mettre en parallèle les estimations d'une enquête et des estimations obtenues de d'autres sources. Voici de grandes questions à se poser lorsqu'on procède à une telle comparaison :

- a. Les populations réelles d'inférence sont-elles les mêmes?
- b. Les questions et les réponses sont-elles d'une formulation identique?
- c. Les questions ont-elles été posées dans des contextes semblables?
- d. Les enquêtes employaient-elles le même mode de collecte de données?
- e. Les enquêtes ont-elles eu lieu au même moment?

Il ressort de ces questions que, dans l'application de cette méthode courante d'évaluation de la qualité d'estimations, on se heurte à des difficultés. On ne se trouve pas à mesurer le seul biais de non-réponse, car les différences peuvent en partie tenir à des différences de mesure ou à des changements réels dans le temps (Federal Committee on Statistical Methodology, 2001). Les différences de mesure l'emportent souvent sur les différences de non-réponse. Ajoutons que les estimations extérieures peuvent aussi être entachées de biais, mais des différences importantes peuvent être révélatrices d'éventuels problèmes. La méthode n'est certes pas tout à fait concluante, mais elle reste une des plus utilisées. Comme la plupart des analystes tâteront au moins un peu de telles comparaisons, cela permet à un organisme de prévoir les inquiétudes que susciteront des données avant même de les diffuser. En règle générale, si on veut rendre les estimations plus comparables, les poids ajustés pour la non-réponse sont utilisés pour effectuer des comparaisons. D'autres ajustements peuvent être apportés afin de rendre les estimations plus comparables. Ainsi, des analystes ont estimé les statistiques d'enquête après avoir constitué des sous-groupes tant dans les données d'enquête que dans les données extérieures de manière à rendre les populations d'inférence identiques ou analogues.

Pour les enquêtes à composition aléatoire, il y a une autre source de comparaison de certaines estimations (Nolin et coll., 2000). Un certain nombre de sociétés recueillent des données au niveau des numéros de téléphone notamment sur le revenu du ménage, l'appartenance de ses membres à diverses catégories d'âge et de sexe, la présence d'enfants, l'instruction des occupants et la taille du logement. Dans l'enquête NHES de 1999, on s'est reporté à des données émanant d'une de ces sociétés pour mettre en parallèle les données de cette enquête et celles de fournisseurs commerciaux sur les répondants. Le taux d'appariement pour l'échantillon des répondants était d'environ 80 %, c'est-à-dire que, pour en gros 80 % des répondants de la NHES, on trouvait des données correspondantes à des fins de comparaison. On relevait toutefois des

différences appréciables entre les estimations de l'enquête et les données du fournisseur commercial en question, ce qui faisait douter de la qualité de ces dernières. On disposait de certaines des variables en question tant pour les répondants que pour les non-répondants, mais à cause de problèmes perçus de qualité, on ne s'en est pas servi dans l'exercice de correction de non-réponse.

Au NCES, on exploite quelques enquêtes extérieures pour les comparaisons. On a déjà employé les estimations de la « Current Population Survey » (Brick, Burke et Lê, 2000) et de la « Survey of Income and Program Participation ». Dans d'autres organismes, on a aussi utilisé comme point de comparaison des données tirées de dossiers administratifs. Il est possible de rapprocher les données recueillies auprès des particuliers de celles de sources extérieures comme les fiches d'hospitalisation et les demandes d'indemnisation d'assurance. Dans les pays ayant un registre national de l'état civil, celui-ci peut servir à une comparaison avec des données d'enquête.

3.4 Lien à établir entre les répondants et la non-réponse

Là où on ne dispose pas de données en provenance ou au sujet des non-répondants, une éventuelle analyse consistera à voir quels répondants « ressemblent » le plus aux non-répondants. Selon le plan d'échantillonnage et la pondération des données, quelques possibilités s'offrent à nous.

3.4.1 Enquêtes à plusieurs volets et à plusieurs poids

Dans une enquête à plusieurs volets, on peut trouver non seulement des unités non déclarantes, mais aussi des non-répondants à certains volets seulement. C'est là un autre niveau de non-réponse qui diffère légèrement de la non-réponse d'unités et de la non-réponse à des questions. L'examen des statistiques d'une enquête en fonction du degré de non-réponse selon ses volets aide à constater le biais de non-réponse qui risque d'apparaître dans des analyses qui se font avec des données relatives à une pluralité de volets.

Ainsi, dans une enquête comme l'ECLS-K, on recueille directement des données d'évaluation auprès des enfants échantillonnés. S'ajoutent des données obtenues des parents, des enseignants des classes ordinaires et de l'enfance en difficulté, des administrateurs d'école et des dossiers scolaires. Les taux de réponse variant selon les volets, on a calculé plusieurs poids. Le choix d'un poids pour l'analyse dépend de ce que celle-ci utilise des données obtenues à un ou plusieurs moments, tout comme des niveaux d'analyse et des sources d'information (un ou plusieurs volets). Ainsi, on a établi un poids pour les enfants avec les données directes d'évaluation au premier cycle de collecte (C-1), de même que pour les enfants avec les trois sources de données (enfants, parents et enseignants des classes ordinaires) à ce même premier cycle (CPT-1). On a créé un troisième poids, un poids permanent, pour les enfants avec toutes les sources de données (au nombre de six) aux premier (trois sources) et deuxième (trois autres sources) cycles de collecte (CPT-P). Ainsi, pour tout enfant avec les six sources de données, on aurait des poids C-1, CPT-1 et CPT-P. Avec trois sources, il y aurait des poids C-1 et CPT-1 et, avec la seule source de l'évaluation des enfants au cycle 1, un poids C-1. Ainsi, les enfants avec un poids CPT-P forment un sous-groupe parmi les enfants ayant un poids CPT-1 qui, eux-mêmes, forment un sous-groupe parmi les enfants avec poids C-1. La différence entre les trois sous-groupes est un autre élément de non-réponse. Ainsi, si on estime la même statistique d'une enquête trois fois, c'est-à-dire avec les poids C-1, CPT-1 et CPT-P, toute différence qu'accusent les estimations peut être mise au compte des différences dues à la non-réponse. Ces différences auraient persisté même après des corrections supplémentaires de non-réponse à chaque pondération et peuvent sans doute être considérées comme des biais introduits par un nouvel élément de non-réponse selon les volets. Il convient de noter que chacun des trois poids a fait l'objet d'une correction distincte pour la non-réponse, ce qui en théorie devrait compenser les différents niveaux de non-réponse (Brick, Burke et Lê, 2000).

Une telle analyse peut se révéler fort utile, car elle se fait comme les chercheurs analysent leurs données. Dans nombre d'analyses, on se reporte aux données de plusieurs sources dans une enquête et, par conséquent, la non-réponse selon les volets risque d'introduire un biais dans les statistiques de cette enquête. Comme dans la plupart des modes d'analyse, l'évaluation se limite aux unités déclarantes et ne

renseigne nullement sur les unités non déclarantes à tous les volets. L'étape suivante de l'analyse considère les répondants pour y aller d'inférences au sujet des non-répondants.

3.4.2 Comparaison entre les répondants « hâtifs » et « tardifs »

Une des grandes hypothèses à la base d'une telle évaluation est que les répondants « tardifs » à une enquête ressemblent plus aux non-répondants que les répondants « hâtifs ». Dans la « Beginning Postsecondary Study », on a modélisé les tendances de la réponse moyenne selon la date de réponse (Wine et coll., 2000). On a divisé les répondants en groupes ayant à peu près la même taille (de sorte que, pour chaque groupe, la réponse moyenne ait en gros la même précision) en fonction de la date d'interview et du type d'établissements. On a regardé les tendances de la population et de sous-populations selon les catégories d'établissements. Dans l'ensemble, cet angle d'examen a permis de constater que les répondants qui se sont ajoutés (répondants tardifs) étaient plus susceptibles d'avoir fréquenté des établissements où les études durent moins de quatre ans et moins susceptibles de fréquenter un établissement postsecondaire au printemps de 1998. L'évaluation s'est faite en utilisant les poids ajustés pour la non-réponse, mais elle est aussi possible en utilisant les poids de base. Limitée aux répondants effectifs, elle se prête néanmoins à une extrapolation des caractéristiques des non-répondants.

3.5 Enquêtes rétrospectives de suivi

Comme nous l'avons mentionné, aucune de ces analyses n'évalue – si ce n'est en utilisant des variables de base de sondage pour une comparaison répondants-non-répondants – si les répondants étaient différents des non-répondants et dans quelle mesure les différences ont introduit un biais dans les diverses estimations. Les enquêtes rétrospectives de suivi permettent aussi une estimation du biais. Elles visent à recueillir des informations pour au moins quelques variables clés ou critiques pour l'ensemble ou un échantillon de non-. On applique des techniques de conversion intensive de non-réponse pour réduire le plus possible la non-réponse de l'échantillon. Le recours à des variables supplémentaires sur les non-répondants donne la possibilité de mieux chiffrer le biais effectif dû à la non-réponse, notamment pour les estimations clés ou les grandes variables de résultats. L'inconvénient est le coût de telles enquêtes de suivi. Il est en outre très important d'obtenir des taux de réponse élevés pour ces études rétrospectives afin qu'elles remplissent leur fonction. Dans l'enquête NELS, on a fait une telle étude de suivi de non-réponse (Spencer et coll., 1990).

3.6 Comparaison des estimations produites avec les poids de base avec celles produites avec les poids ajustés pour la non-réponse

Dans l'établissement d'un poids ajusté pour la non-réponse, il faut entre autres relever les caractéristiques les plus reliées à la non-réponse. On recourt à des analyses multivariées variables pour dégager des sous-groupes en fonction des différences de propension à répondre à l'enquête. L'hypothèse est que, dans ces sous-groupes, les répondants et les non-répondants auront des réponses semblables et qu'il existera de grandes différences entre sous-groupes. On délimite des cellules selon des caractéristiques communes aux répondants et aux non-répondants. Dans chaque cellule généralement définie par plusieurs variables, on applique un facteur d'ajustement de poids aux répondants en compensation de la non-réponse, le but étant d'éliminer ou de réduire le biais de non-réponse. On peut procéder à une analyse de la propension à répondre à l'aide d'un algorithme de recherche catégorique appelé « Chi-Square Automatic Interaction Detection » (CHAID). On peut diviser un ensemble de données en cellules de sorte que toutes les unités d'une cellule aient les mêmes probabilités de réponse selon ce que déterminera l'analyse.

Une autre façon d'évaluer l'effet des ajustements de non-réponse sur différentes estimations d'une enquête est de regarder les estimations obtenues en utilisant le poids de base à celles obtenues avec le poids ajusté. Si les différences sont grandes, il est possible que la correction apportée ait bel et bien eu pour effet de réduire le biais dont elles sont entachées. S'il n'y a pas de différences, il est par ailleurs possible que l'échantillon initial de répondants n'ait pas été très différent des non-répondants et qu'il n'y ait pas eu de biais outre mesure au départ, mais il se peut aussi que les caractéristiques ayant servi à la délimitation des cellules n'aient pas permis de bien prévoir la propension à répondre. Somme toute, cette méthode est utile à

une évaluation de l'incidence des corrections de non-réponse sur les estimations, mais sans nécessairement nous éclairer sur le degré de biais qu'accusent les estimations d'une enquête.

3.7 Autres méthodes

Dans une enquête longitudinale, on peut, une fois recueillies les données des répondants dans l'année de référence, comparer les non-répondants aux répondants aux cycles ultérieurs en utilisant plus que les seules données d'une base de sondage. Bien sûr, cela ne nous renseigne pas sur la non-réponse initiale des unités de l'échantillon, mais on peut obtenir des indications sur le biais d'érosion qui est peut-être lié à la non-réponse supplémentaire aux cycles subséquents de collecte.

D'autres possibilités existent. Les statisticiens d'autres organismes ont examiné les interviews partielles ou interrompues par rapport aux interviews complètes en supposant que les enquêtés qui risquent de ne pas achever l'interview ressembleront davantage aux non-répondants. De même, on a affirmé la possibilité dans d'autres études que les répondants en conversion de refus ou plus réfractaires au départ dans une enquête se comparent plus à des non-répondants qu'à des répondants.

4. ÉVALUATION DU BIAIS ESTIMÉ

Il existe diverses façons d'évaluer le biais. Si on dégage sa valeur absolue, on ne se trouve pas bien renseigné pour autant au sujet de son incidence sur les estimations. Dans les enquêtes du NCES, on a trouvé quelques façons d'évaluer le biais estimé.

On détermine si le biais est différent de zéro : Si l'intervalle de confiance dont on encadre un biais contient la valeur zéro, on peut juger que celui-ci n'est pas significatif. C'est une technique qu'on a appliquée au moment de comparer, par exemple, les statistiques d'enquête aux valeurs de population tirées de la base de sondage. Le biais est considéré comme la différence entre une statistique d'enquête et une valeur de population et, si l'intervalle de confiance contient la valeur zéro, on est à même de déterminer s'il existe un biais quelconque.

On compare l'importance du biais par rapport aux valeurs des statistiques de l'enquête : Une façon simple d'examiner le biais est de le comparer à la statistique de l'enquête. En calculant un tel biais relatif, on permet des comparaisons sur la diversité des estimations de l'enquête. Cela ne nous dit rien cependant du biais par rapport à la confiance que l'on peut avoir dans une statistique selon l'erreur-type. Dans des enquêtes, on dégage cependant une valeur moyenne de « biais relatif », c'est-à-dire une moyenne de valeurs multiples de biais relatif.

On compare également l'importance du biais par rapport à l'écart-type : Il est également possible de comparer le biais estimé à l'écart-type de la statistique de l'enquête. Par l'écart-type d'une estimation, on peut souvent relever d'importantes différences de fond.

Finalement, on compare l'importance du biais par rapport à l'erreur-type : L'évaluation du biais estimé peut aussi se faire par rapport à l'erreur-type. L'erreur quadratique moyenne peut ainsi s'exprimer :

$$\text{Erreur quadratique moyenne} = (\text{biais})^2 + \text{variance}$$

Ainsi, si le biais est important par rapport à l'erreur-type, il contribue le plus à l'erreur quadratique moyenne. Dans de grands échantillons, le biais sera souvent élevé par rapport à l'erreur-type.

CONCLUSIONS

Dans des enquêtes, il est bon d'obtenir de hauts taux de réponse, qui ne sont toutefois pas synonymes de faibles biais là où répondants et non-répondants sont très différents, mais il demeure que, avec des taux de

réponse plus faibles, on se trouve à amplifier d'autant les effets de la différence entre réponse et non-réponse qui contribue à ces biais. Après la collecte des données d'une enquête, nos analyses aident à en établir la qualité et les éléments de vulnérabilité et à améliorer la collecte aux cycles ultérieurs d'études longitudinales et aux reprises d'enquêtes transversales. Les modes d'analyse abondent. Il est possible d'adapter les analyses aux caractéristiques des enquêtes. Par comparaison avec une base de sondage ou par recours aux données d'une enquête rétrospective de suivi, on peut estimer le biais dû à la non-réponse d'unités. On peut également estimer le surcroît de biais introduit en se reportant à un sous-ensemble de cas avec des données complètes, surtout si on a pris soin de bien calculer les poids pour les sous-groupes retenus. Il n'est pas toujours possible de mesurer le biais avec exactitude, mais les analyses de biais de non-réponse font partie intégrante de l'évaluation générale de la qualité des données.

BIBLIOGRAPHIE

- Brick, J. M., J. Burke, et T. Lê (2000), "Analysis of Nonresponse Bias in the Base Year, Early Childhood Longitudinal Study: Kindergarten Class of 1998-99", rapport non publié, Washington, DC: National Center For Education Statistics.
- Federal Committee on Statistical Methodology (2001), *Measuring and Reporting Sources of Errors in Surveys*, Washington, DC: U.S. Office of Management and Budget (Statistical Working Paper 31), pp. 6-1 – 6-13.
- Green, P., S. Myers, C. Veldman, S. Pedlow, et P. R. Knepper (1999), "Nonresponse Bias Analysis", *Baccalaureate and Beyond Longitudinal Study: 1993/97 Second Follow-Up Methodology Report*, pp. 62 – 78, (NCES 1999-159).
- Monaco, D., S. Salvucci, F. Zhang, et M. Hu (1997), *An Analysis of Total Nonresponse in the 1993-94 Schools and Staffing Survey*, U.S. Department of Education, National Center for Education Statistics (NCES 98-243).
- Nolin, M. J., J. Montaquila, P. Nicchitta, K. Kim, B. Kleiner, J. Lennon, C. Chapman, S. Creighton, et S. Bielick (2000), "A Study of Nonresponse Bias in the NHES: 99", *National Household Education Survey: 1999 Methodology Report*, pp. 122 – 143, (NCES 2000-078).
- Spencer, B. D., M. R. Frankel, S. J. Ingels, K. A. Rasinski, R. Tourangeau, et J. A. Owings (1990), "School Nonresponse Analysis", *National Education Longitudinal Study of 1988, Base Year Sample Design Report*, pp. 33 – 48 (NCES 90-463).
- Wine, J. S., R.W. Whitmore, R. E. Heuer, M. Biber, D. J. Pratt, et C. D. Carroll (2000), "Measures of Bias", *Beginning Postsecondary Students Longitudinal Study First Follow-up 1996-98 (BPS: 96/98)*, pp. 6-41 – 6-64, (NCES 2000-157).