

L'estimation à partir d'enquêtes non probabilistes et la question de la validité externe

Jill A. Dever et Richard Valliant¹

Résumé

Les enquêtes probabilistes, c'est-à-dire celles réalisées à partir d'échantillons sélectionnés selon un mécanisme aléatoire connu, sont considérées par plusieurs comme étant la norme par excellence comparativement aux enquêtes réalisées à partir d'échantillons non probabilistes. La théorie de l'échantillonnage probabiliste a vu le jour au début des années 1930 et sert encore aujourd'hui à prouver le bien-fondé d'estimer les valeurs d'une quelconque population à partir des données de ces enquêtes. Les études sur les échantillons non probabilistes, quant à elles, ont retenu davantage l'attention ces dernières années, mais elles ne sont pas nouvelles. Vantées comme étant moins coûteuses, plus rapides (voire meilleures) que les études avec plan probabiliste, ces enquêtes s'appuient, pour recruter les participants, sur diverses méthodes « de terrain » (p. ex., enquête en ligne à participation volontaire). Mais, quel type d'enquêtes est le meilleur?

Le présent article est le premier d'une série portant sur la recherche d'un cadre de la qualité permettant de mesurer toutes les enquêtes, probabilistes ou non, en les plaçant davantage sur un pied d'égalité. D'abord, nous donnons un aperçu de quelques cadres utilisés à l'heure actuelle, en notant que le terme « meilleur » se rapporte presque toujours à la notion « d'adaptation de l'enquête à son but poursuivi » (*fit for purpose*). Ensuite, nous nous concentrons sur la question de la validité, surtout la validité externe, lorsque des estimations pour la population sont souhaitées. Nous passons en revue les techniques d'estimation appliquées jusqu'à présent aux enquêtes non probabilistes, ainsi que quelques études comparant ces estimations à celles obtenues avec un échantillon probabiliste. Enfin, nous décrivons les prochaines étapes de recherche et concluons par quelques remarques.

Mots-clés : Enquêtes avec échantillon non probabiliste; adaptation de l'enquête à son but poursuivi; cadre de la qualité; validité externe.

1. Introduction

Innovation! Un mot que l'on ne cesse d'entendre et d'utiliser de nos jours. De nombreux livres traitent du sujet et on se demande comment nous pourrions mieux faire les choses. Par exemple, dans un ouvrage récent, Bryce (2014) associe l'innovation (et donc la définition de « meilleur ») aux caractéristiques « plus petit, plus rapide, plus léger, plus dense, moins cher ». Ce même sentiment s'applique également à la recherche en sondage. Toutefois, comment pourrions-nous définir le mot « meilleur » dans le contexte des enquêtes?

Holt (2007), par exemple, définit « meilleur » en utilisant des termes tels que plus vaste, plus profond, plus rapide et moins cher. Selon lui, les « pressions incessantes » exercées sur les organismes gouvernementaux afin qu'ils produisent des statistiques officielles de haute qualité à un rythme rapide ont un côté positif, en ce sens qu'il s'agit d'un moyen d'accroître l'innovation dans le domaine de la recherche en sondage. Cette opinion a également été soutenue par Constance F. Citro, directrice du Committee on National Statistics², lors de son discours comme gagnante du prix Waksberg lors du Symposium 2014 de Statistique Canada. Les pressions de ce genre ne s'exercent pas seulement sur ceux qui produisent des statistiques officielles, mais aussi sur tout le monde de la recherche.

Les décideurs, les professionnels de la santé, les organismes gouvernementaux et les membres du grand public souhaitent tous de l'information rapide. Par exemple, le U.S. Center for Disease Control réalise une enquête auprès

¹Jill A. Dever, RTI International, 701 13th St NW, Suite 750, Washington, DC 20005-3967, jdever@rti.org; Richard Valliant, Institute for Social Research à l'Université du Michigan, Joint Program in Survey Methodology à l'Université du Maryland, 1218 LeFrak Hall, College Park, MD 20742, rvallian@umd.edu.

² <http://sites.nationalacademies.org/dbasse/cnstat/index.htm>.

de la population générale sur les oppositions possibles à la vaccination antigrippale tôt dans la saison afin d'appuyer les campagnes d'éducation en matière de santé menées chaque année en décembre (Srivastav et coll., 2014). Rien n'est plus critique que d'obtenir l'information rapidement lorsque la nouvelle d'une catastrophe naturelle ou d'une maladie infectieuse arrive sur les ondes. Tout récemment, des enquêtes ont été menées aux États-Unis pour évaluer ce que le public savait sur la probabilité d'être infecté par le virus Ebola (voir, p. ex., Dwyer, 2014), de même que le degré de préparation des professionnels de la santé en vue de traiter les cas possibles (voir, p. ex., Garman, 2014). Cependant, la diminution de la participation et les besoins sans cesse croissants en ressources (coût et temps) associés aux échantillons probabilistes amènent certains à mettre en doute l'utilité des plans de sondage classiques.

L'enquête avec échantillon non probabiliste est l'une des réponses offertes par les chercheurs à cette demande d'innovation. Comme l'a fait remarquer Smith (1976), l'utilisation d'échantillons non probabilistes existe depuis longtemps. Vantées comme étant moins coûteuses, plus rapides (voire même meilleures) que les enquêtes utilisant des échantillons probabilistes, ces enquêtes s'appuient, pour recruter les participants, sur diverses méthodes « de terrain » (p. ex., enquête en ligne à panel volontaire). Un cas extrême, vu en pratique, d'un « bon » échantillon non probabiliste est un échantillon d'utilisateurs de la console de jeu Xbox à partir duquel on a prédit correctement le résultat de l'élection présidentielle américaine de 2012 (Wang et coll. 2015). Des résultats favorables tels que ceux-là portent certaines personnes à croire que l'échantillonnage probabiliste pourrait être mis de côté entièrement. Cependant, ne sacrifions-nous pas ainsi la qualité ou la généralisabilité (c.-à-d. la validité externe) dans cette quête de données plus rapides et moins coûteuses? À notre avis, la réponse à cette question est compliquée (« cela dépend » si vous voulez). Selon les Lignes directrices concernant la qualité (2009) de Statistique Canada, « la qualité est relative, et non absolue ».

La réponse dépend des diverses mesures de la qualité existantes qui sont passées en revue à la première section de notre article. Ensuite, nous discutons de deux dimensions importantes dans le contexte du paradigme de « l'adaptation de l'enquête à son but poursuivi » et mentionnons certains facteurs de complication dans le cas des enquêtes non probabilistes (section 3). À la section 4, nous résumons les techniques utilisées à l'heure actuelle pour produire des estimations découlant d'enquêtes non probabilistes afin de cerner les enjeux. S'il y a lieu, les résultats mitigés obtenus au moyen de ces méthodes comparativement à leurs analogues probabilistes sont mentionnés (section 5). Nous soulignons quelques questions qu'il faudra aborder dans le cadre de futurs travaux de recherche (section 6) et concluons en résumant nos idées.

2. Qualité des enquêtes

Certaines personnes éprouvent de la difficulté à définir l'art, mais elles savent le reconnaître. Dans le passé, on pouvait en dire autant de la qualité des enquêtes. Au départ, certains chercheurs ont essayé de trouver une mesure de la qualité globale d'une enquête permettant de faire facilement des comparaisons entre les enquêtes. Toutefois, cette tâche s'est avérée difficile. Prenons, par exemple, le taux de réponse. Groves et Peytcheva (2008) ont montré que le taux de réponse n'est pas nécessairement une bonne mesure indirecte de la qualité des données, parce qu'il n'existe pas systématiquement une forte corrélation entre un faible taux et une plus grande présence d'un biais de non-réponse dans les données.

Selon les Lignes directrices concernant la qualité de Statistique Canada (2009), « ces 20 dernières années, les organismes statistiques sont parvenus à un consensus : la *qualité* de l'information statistique est multidimensionnelle. » (Voir aussi Biemer et Lyberg, 2003). Cependant, les dimensions de la qualité ne sont pas toutes égales. Entre alors en jeu l'expression « adaptation de l'enquête à son but poursuivi » (de l'anglais *fit for purpose* et réduite à l'expression « adaptation au but poursuivi » dans le reste du texte) dans le vocabulaire de la recherche en sondage pour l'élaboration de plans d'étude (Baker et coll., 2013). Cette expression existe depuis longtemps pour décrire la qualité (ou le manque de celle-ci) dans divers domaines. L'adaptation au but poursuivi signifie qu'une chose est « suffisante pour accomplir la tâche pour laquelle elle a été conçue » [traduction] (<http://www.macmillandictionaryblog.com/>) et « est appropriée, et satisfait à la norme nécessaire, à son usage prévu » [traduction] (<http://en.wiktionary.org/>). L'approche de l'adaptation au but poursuivi détermine (ou du moins devrait déterminer) laquelle d'une série de dimensions de la qualité est la plus importante. Par exemple, dans le cas d'une enquête sur la flambée d'une maladie infectieuse, on pourrait donner la priorité à l'estimation en temps réel plutôt qu'à toute autre dimension.

Dans les deux sections qui suivent, nous examinons deux cadres de la qualité décrits dans la littérature pour définir les dimensions de l'adaptation au but poursuivi. Jusqu'à présent, ces cadres ont été appliqués principalement à des enquêtes probabilistes. La dernière section met brièvement en relief certains défis que pose la définition d'un cadre comparable pour les enquêtes non probabilistes.

2.1 Cadre d'assurance de la qualité de Statistique Canada

Le cadre d'assurance de la qualité de Statistique Canada (2002) mentionne six mesures de la qualité interdépendantes :

- Pertinence – les données satisfont aux besoins spécifiés par les clients (p. ex., organisme de financement, chercheurs et membres du public).
- Exactitude – les données peuvent être utilisées pour mesurer correctement la population cible, comme il est prévu, avec des niveaux acceptables de précision et d'exactitude.
- Actualité – les données ont trait à la période en question et sont disponibles au moment où elles sont nécessaires.
- Accessibilité – les données sont faciles à obtenir en utilisant un moyen de collecte des données approprié à un coût raisonnable.
- Intelligibilité – les données et le plan de sondage sont facilement compris.
- Cohérence – les données concordent avec d'autres sources de données.

2.2 Cadre de la qualité basé sur l'erreur d'enquête totale

L'erreur d'enquête totale (EET) englobe l'ensemble des travaux de recherche orientés vers la détection et le traitement d'une foule d'erreurs qui peuvent survenir durant une enquête, y compris les erreurs d'échantillonnage et les défis associés à la non-réponse. Biemer (2010) discute d'un cadre de la qualité vu à travers la lentille de l'EET qui inclut les six dimensions mentionnées par Statistique Canada (section 2.1), ainsi que trois autres :

- Crédibilité – les données sont considérées comme étant fiables.
- Comparabilité – les méthodes appliquées dans des études similaires sont compatibles.
- Complétude – les données répondant aux objectifs d'analyse sont obtenues sans « imposer de fardeau excessif aux répondants ».

2.3 Cadres de la qualité pour les enquêtes non probabilistes

Baker et coll. (2013), entre autres, discutent de quelques défis à relever pour évaluer la qualité des enquêtes non probabilistes. Par exemple, dans le cas de nombreuses enquêtes non probabilistes, il n'existe pas de base de sondage, laquelle ne peut donc pas être examinée par les chercheurs. Voici certaines réflexions supplémentaires au sujet de quelques dimensions énumérées plus haut :

- Exactitude/comparabilité – les estimations provenant d'enquêtes non probabilistes ont parfois concordé avec des estimations comparables provenant d'autres sources mais pas toujours (voir la section 4 pour cette discussion).
- Cohérence – les incohérences par rapport à d'autres sources de données est un problème qui se pose aussi bien pour les enquêtes probabilistes que non probabilistes et sont difficiles à expliquer.
- Crédibilité – la fiabilité des enquêtes en général est mise en doute si la validité des données n'est pas prouvée ou si le niveau de non-participation (s'il est connu) est jugé excessif par les chercheurs.

Des mots tels qu'actualité, accessibilité et complétude sont parfois utilisés pour justifier le choix d'une enquête non probabiliste plutôt qu'une enquête fondée sur une méthode d'échantillonnage aléatoire définie.

3. Généralisation des estimations tirées d'enquêtes

De plus en plus souvent, les plans d'étude (c.-à-d. une combinaison des conditions essentielles d'enquête, telles que le plan de sondage et le mode de collecte des données [Hansen et coll., 1961]) sont choisis de manière à maximiser l'adaptation de l'enquête à son but poursuivi. Deux dimensions de la qualité importantes associées à cette dernière, et sur lesquelles se concentre le présent article, sont la validité et l'exactitude.

3.1 Validité

La validité se décline en deux catégories et a un effet direct sur le choix du plan de sondage. L'expression *validité interne* (voir, p. ex., <http://www.edpsycinteractive.org/>) laisse entendre que toutes les variables d'intérêt ont été mesurées (et bien mesurées) sur les participants à l'étude et qu'elles sont disponibles pour les analyses. Cette expression se retrouve à travers toute la littérature sur les études, entre autres, celle sur les essais cliniques. L'expression *validité externe* (voir, p. ex., <http://www.socialresearchmethods.net/>) indique quant à elle que les estimations fondées sur l'étude peuvent être généralisées au-delà de l'échantillon, disons à la population cible, et qu'en fait, elles peuvent être reproduites.

Les chercheurs qui se servent d'enquêtes non probabilistes font tout spécialement l'objet de critiques concernant le manque de validité externe. L'une des raisons est que nombre de ces études ne possèdent pas de base de sondage et qu'elles manquent donc d'un lien discernable avec la population cible (voir, p. ex., Baker et coll., 2013). Les participants aux enquêtes non probabilistes peuvent avoir été « attrapés » par coïncidence et pourraient ne pas être accessibles si l'enquête était de nouveau effectuée. Un lien allant de l'échantillon non probabiliste à la population offre l'espoir qu'un modèle commun puisse être utilisé pour prédire les valeurs des variables de l'analyse pour les deux ensembles d'unités (p. ex., voir Valliant et coll., 2000). Cependant, il peut être difficile, voire impossible, de diagnostiquer si un modèle est vérifié pour les deux ensembles. D'où la justification que l'« hypothèse de reproductibilité » pourrait être faible.

3.2 Exactitude

L'exactitude, une fonction du biais (au carré) et de la précision, est la différence entre l'estimation et le paramètre d'intérêt prévu (réel) que l'on veut mesurer. Dans le monde des enquêtes, une estimation est considérée comme étant exacte si l'erreur quadratique moyenne (= biais² + variance) est faible. Nous discutons brièvement de la précision des enquêtes non probabilistes à la section suivante et remettons son examen plus approfondi à une autre occasion. Cela nous laisse le biais.

Pour décrire trois composantes du biais dans les enquêtes, Valliant et Dever (2011) commencent par définir trois populations : 1) la *population cible* (d'intérêt) pour l'étude, 2) la *population potentiellement couverte*, disponible en vertu des conditions essentielles d'enquête et 3) la *population réellement couverte*, qui est la proportion de la population cible dont les membres sont sélectionnés par l'étude en appliquant les conditions essentielles d'enquête. Par exemple, considérons une enquête en ligne à participation volontaire pour une étude sur le tabagisme. La population cible pourrait être définie comme étant les adultes de 18 à 29 ans qui fument des cigarettes à l'heure actuelle. La population potentiellement couverte comprendrait les personnes admissibles à l'étude qui ont accès à Internet et qui visitent les sites où a lieu le recrutement pour l'étude. La population réellement couverte correspondrait quant à elle au sous-ensemble de la population potentiellement couverte qui est représenté par les participants à l'étude.

Ayant spécifié les trois populations, les auteurs définissent ensuite les trois composantes du biais suivantes :

- biais de couverture – différence entre la population cible et la population potentiellement couverte;
- biais de non-réponse / non-participation – différence entre les populations potentiellement et réellement couvertes;
- biais de sélection – différence entre la population cible et la population réellement couverte.

La littérature sur les enquêtes est remplie d'études conçues pour évaluer l'ampleur du biais de couverture associé à certaines bases de sondage, comme les listes de numéros de téléphone fixe aux États-Unis (voir, p. ex., Christian et coll., 2010) et avec certains modes de collecte des données, comme Internet (voir, p. ex., Dever et coll., 2008). Les références à l'ajustement des poids fondés sur les probabilités de sélection pour corriger le biais de non-réponse sont nombreuses (voir, p. ex., Kott, 2006). Le terme biais de sélection se retrouve partout dans la littérature sur les enquêtes non probabilistes (voir, p. ex., Lee et Valliant, 2009) puisque, dans le cas de nombreuses études, les caractéristiques des membres de l'échantillon (potentiellement) non participants, ou même le pourcentage de non-répondants, sont inconnues.

Nombre de chercheurs se servant d'enquêtes avec échantillon probabiliste s'appuient sur la théorie assistée par le plan de sondage - ou assistée par un modèle - ou sur des modèles pour limiter les erreurs dans les données afin de prouver qu'il y a absence de biais avec des estimations sous le plan de sondage (voir, p. ex., Särndal et coll., 1992). Par exemple, le calage des poids peut être utilisé pour minimiser le biais associé à la non-réponse et à la couverture (Kott, 2006). À la section suivante, nous résumons les techniques d'ajustement pour enquêtes non probabilistes qui visent au moins à corriger le biais de sélection, ainsi que quelques études réalisées pour évaluer leur utilité.

4. Méthodes d'estimation pour les enquêtes non probabilistes

Dans la présente section, nous résumons certaines des méthodes d'ajustement et d'estimation utilisées jusqu'à présent pour maximiser l'utilité des données obtenues au moyen d'enquêtes non probabilistes. Par souci de commodité, nous classons ces méthodes en deux catégories, à savoir les méthodes fondées sur un modèle et les pseudo-méthodes fondées sur un plan de sondage.

4.1 Méthodes fondées sur un modèle

La première méthode est celle de l'*approche fondée sur un modèle de la population*. Ici, les chercheurs utilisent les données d'enquête dans un modèle établi pour la population d'intérêt. Les hypothèses que nécessite cette approche, comme il est exposé dans de nombreux manuels de statistiques mathématiques, sont : qu'un échantillon soit aléatoirement tiré de la population (c.-à-d. que les données dans l'échantillon et dans la population suivent le même modèle, ce qui est vrai si les données ont été ajoutées ou non à l'échantillon de façon aléatoire [voir, p. ex., Valliant et coll., 2013]) et que toute composante aléatoire possible soit mesurée et comprise dans le modèle. Bien que des estimations pour la population soient produites, la vérification des hypothèses n'est pas toujours discutée. Les exemples pertinents tirés de l'approche fondée sur un modèle de la population sont trop nombreux pour être tous survolés dans ce bref exposé. La théorie qui sous-tend cette approche est bien établie (Valliant et coll. 2000). En revanche, les méthodes permettant de vérifier qu'un modèle basé sur un échantillon non probabiliste représente bien la population au complet ne le sont pas.

Une autre méthode fondée sur un modèle est connue sous le nom de *modèle de Heckman* (1979). Elle a été élaborée pour la modélisation économique au moyen de données observationnelles et le mécanisme d'inclusion dans l'étude est estimé en utilisant un modèle initial. La correction résultante du biais de sélection est ensuite incluse dans un modèle d'estimation subséquent. Cette technique s'appuie fortement sur l'hypothèse d'une distribution normale sous-jacente pour estimer le biais de sélection. Comme dans le cas de la première méthode, l'hypothèse des enregistrements sélectionnés au hasard pour l'échantillon est importante.

La dernière approche fondée sur un modèle que nous commentons brièvement est la *modélisation bayésienne*. Comme dans le cas de la modélisation bayésienne classique ou de la modélisation d'une superpopulation, les unités comprises dans la population et dans l'échantillon doivent suivre le même modèle. L'établissement de lois a priori

pour les paramètres du modèle peut mener à des classes d'estimateurs plus riches qu'en l'absence de lois a priori. À l'heure actuelle, cette approche n'est pas très répandue, mais l'intérêt qu'elle suscite augmente (voir, p. ex., Little, 2004; Roshwalb et coll., 2012).

Avant de conclure cette brève discussion sur les méthodes fondées sur un modèle pour les enquêtes non probabilistes, nous devons aborder la question des données sélectionnées au hasard. Certaines enquêtes, dont les mécanismes de sélection aléatoire sont indéfinissables, sont accompagnées d'allégations selon lesquelles le hasard expliquerait la « non-sélection » des non-participants, et ce, parce que la distribution démographique des participants concorde avec celle de la population d'intérêt. Cependant, comme le soulignent Dever et coll. (2008), les données démographiques ne rendent pas nécessairement possible d'établir un bon ajustement des poids permettant de bien couvrir les données manquantes en question. Des tests pour déterminer si le hasard explique réellement la « non-sélection » des non-participants existent, mais à l'heure actuelle, uniquement pour des enquêtes avec échantillon aléatoire (Pfeffermann et Sikov, 2011).

4.2 Pseudo-estimation fondée sur un plan de sondage

L'hypothèse selon laquelle le facteur aléatoire déterminerait quelles données seront manquantes est également utilisée dans le cas de ce que nous appelons la pseudo-estimation fondée sur un plan de sondage. Ici, les chercheurs utilisent des données provenant d'une enquête pour laquelle il n'existe pas de mécanisme d'échantillonnage aléatoire « défini » et les traitent comme si elles avaient été recueillies à partir d'un *échantillon aléatoire simple (stratifié)* (voir, p. ex., Thompson, 2002). Autrement dit, l'enquête non probabiliste est reclassifiée comme possédant un plan d'échantillonnage à un degré simpliste alors que, dans de nombreux cas, il n'existait pas de base de sondage à partir de laquelle tirer un tel échantillon. Les poids de base, dont la valeur est généralement fixée à un, sont calés sur les totaux de population connus (Loosveldt et Sonck 2008). Les poids d'analyse calés sont ensuite utilisés avec les données et les estimateurs d'Horvitz-Thompson classiques (1952) pour calculer les statistiques de la population (Deville et Särndal, 1992). Cette méthode peut aussi être considérée comme fondée sur un modèle au sens où les variables de calage impliquent un modèle qui doit être vérifié pour les parties de la population à l'intérieur comme à l'extérieur de l'échantillon.

La technique suivante est connue sous le nom de méthode d'*ajustement par le score de propension* et est fondée sur les travaux de Rosenbaum et Rubin (1983) concernant les études observationnelles. Dans sa version adaptée pour la recherche en sondage, un échantillon non probabiliste est combiné à un échantillon probabiliste (de référence) pour estimer la propension à répondre à une enquête non probabiliste. L'inverse de la propension est utilisé directement pour former les poids d'analyse ou indirectement pour former des post-strates (voir, p. ex., Lee et Valliant 2009). De nouveau, on peut caler les poids d'analyse sur les totaux de population connus, puis utiliser les poids d'analyse calés pour créer des estimations de type Horvitz-Thompson.

5. Études comparatives

Jusqu'à présent, les publications ayant trait à l'utilité de l'ajustement dans le cas des enquêtes avec échantillon non probabiliste ont été peu abondantes, mais la situation est en train de changer. Par exemple, Tourangeau et coll. (2013) ont examiné huit panels en ligne volontaires en utilisant des techniques d'ajustement des poids et ont indiqué que les résultats étaient variables. En général, les ajustements n'éliminaient qu'une partie seulement du biais de sélection et, dans certains cas, accroissaient en fait le biais comparativement aux estimations non ajustées. Les avantages de l'ajustement variaient aussi fortement selon le type de variable d'analyse. Yeager et coll. (2011) ont obtenu des résultats tout aussi mitigés en comparant une enquête utilisant la méthode de composition aléatoire (CA) à une enquête en ligne non probabiliste.

Valliant et Dever (2011) résument les hypothèses nécessaires pour adapter la méthode d'ajustement par le score de propension de Rosenbaum et Rubin (1983) au domaine des enquêtes. Plus précisément, les deux enquêtes doivent inclure des participants choisis aléatoirement - parmi lesquels le hasard expliquerait la « non-sélection » des non-participants - et les échantillons des deux enquêtes ne peuvent se chevaucher. Ils constatent que la deuxième hypothèse est clairement violée dans la plupart des situations d'enquête de référence non probabiliste, sinon toutes.

Les auteurs notent que l'utilisation des poids d'analyse pour l'enquête de référence était nécessaire dans l'estimation des scores de propension afin de minimiser le biais de sélection. D'autres travaux doivent être effectués afin de déterminer quel est le meilleur poids d'entrée qu'il convient d'utiliser pour l'échantillon non probabiliste dans les méthodes fondées sur le score de propension.

6. Recherche future

D'autres études comparatives publiées donnent à penser que les résultats concernant la validité externe des estimations fondées sur un échantillon non probabiliste sont variables pour une même enquête, ainsi que lorsqu'on compare diverses enquêtes. Les travaux de recherche futurs pourraient inclure la lourde tâche consistant à déterminer quand les estimations non probabilistes donnent de bons résultats. Selon Valliant (2013), les modèles pourraient s'avérer un outil important en vue de trouver la réponse à ce mystère. En outre, la réponse doit inclure un cadre de la qualité s'appliquant à toutes les enquêtes conçues en vue d'atteindre l'objectif d'adaptation de l'enquête à son but poursuivi.

En particulier, les spécialistes de la recherche en sondage qui souhaitent définir la place appropriée (s'il en existe une) des enquêtes avec échantillon sans mécanisme d'échantillonnage aléatoire défini (c.-à-d. sans échantillon probabiliste) ont beaucoup de travail à accomplir. Nous définissons les objectifs de la recherche future sous forme d'une série de questions et espérons ajouter bientôt la réponse à certaines d'entre elles à nos connaissances de base.

- 1) Existe-t-il des dimensions de la qualité pour comparer de manière appropriée les enquêtes probabilistes et non probabilistes, ainsi que les enquêtes non probabilistes entre elles?
- 2) Existe-t-il des critères pour réétiqueter une enquête probabiliste comme étant non probabiliste en se basant, par exemple, sur les faibles taux de réponse, et donc passer à ce nouveau cadre de la qualité?
- 3) À l'intérieur du cadre de la qualité, comment pourrions-nous définir un ensemble de critères normalisés pour « accepter » les statistiques produites au moyen d'une enquête non probabiliste comme étant des estimations pour la population?

Les mesures équilibrées qui ont été proposées pour l'étude de la non-réponse (Särndal et Lundquist 2014), ayant joué un rôle clé dans l'optimalité dans l'approche fondée sur un modèle (Royall 1992), pourraient vraiment être d'intérêt en ce qui concerne la question (3).

7. Conclusion

L'innovation a permis à la recherche en sondage de faire de grands progrès et elle continuera de le faire dans l'avenir. Les enquêtes non probabilistes sont considérées par certains comme un moyen novateur de réagir aux pressions croissantes en vue de produire des données moins coûteuses, plus rapidement, surtout compte tenu du fait que la non-réponse aux enquêtes probabilistes ne cesse d'augmenter. Cependant, la question à savoir si les enquêtes non probabilistes sont simplement un remède miracle plutôt qu'une solution réelle suscite de nombreux débats. Le présent article est le premier d'une série ayant pour but de définir des solutions et un cadre de la qualité connexe pour toutes les enquêtes dans l'optique de l'adaptation d'une étude à son but poursuivi. Ici, nous avons présenté seulement quelques résultats mitigés qui se dégagent de la recherche sur les enquêtes non probabilistes jusqu'à présent, afin d'encadrer la discussion. Beaucoup de travail reste à faire, mais avons-nous déjà reculé devant un défi?

Bibliographie

Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A. et coll. (2013), « Summary report of the AAPOR task force on non-probability sampling », *Journal of Survey Statistics and Methodology*, 1, p. 90 à 143.

- Biemer, P. P. (2010), « Total Survey Error: Design, Implementation, and Evaluation », *Public Opinion Quarterly*, 74(5), p. 817 à 848.
- Biemer, P. P. et Lyberg, L. E. (2003), *Introduction to Survey Quality*, New Jersey: John Wiley & Sons, Inc.
- Bryce, R. (2014), *Smaller Faster Lighter Denser Cheaper: How Innovation Keeps Proving the Catastrophists Wrong*, New York: PublicAffairs™.
- Christian, L., Keeter, S., Purcell, K. et Smith, A. (2010), « Assessing the Cell Phone Challenge », Pew Research Center, rapport disponible à l'adresse <http://www.pewresearch.org/2010/05/20/assessing-the-cell-phone-challenge/>.
- Dever, J.A., Rafferty, A. et Valliant, R. (2008), « Internet Surveys: Can Statistical Adjustments Eliminate Coverage Bias? » *Survey Research Methods*, 2(2), p. 47 à 62.
- Deville, J.-C. et Särndal, C.-E. (1992), « Calibration estimators in survey sampling », *Journal of the American Statistical Association*, 87, p. 376 à 382.
- Dwyer, M. (2014), « Poll finds many in U.S. lack knowledge about Ebola and its transmission », communiqué de presse de la Harvard School of Public Health (21 août 2014), disponible à l'adresse <http://www.hsph.harvard.edu/news/press-releases/poll-finds-many-in-us-lack-knowledge-about-ebola/>.
- Garman, L. (2014), « APIC [Association for Professionals in Infection Control and Epidemiology] Ebola Readiness Poll: Results of an online poll of infection preventionists », disponible à l'adresse http://www.apic.org/Resource/TinyMceFileManager/Topic-specific/Ebola_Readiness_Poll_Results_FINAL.pdf.
- Groves, R. et Peytcheva, E. (2008), « The impact of nonresponse rates on nonresponse bias: A meta-analysis », *Public Opinion Quarterly*, 72 (2), p. 167 à 189.
- Hansen, M.H., Hurwitz, W.N. et Bershad, M.A. (1961), « Measurement Errors in Censuses and Surveys », *Bulletin of the International Statistical Institute*, 38(2), p. 359 à 374.
- Heckman, J.J. (1979), « Sample Selection Bias as a Specification Error », *Econometrica*, 47, p. 153 à 62.
- Horvitz, D.G. et Thompson, D.J. (1952), « A generalization of sampling without replacement from a finite universe », *Journal of the American Statistical Association*, 47, p. 663 à 685.
- Holt, D.T. (2007), « The Official Statistics Olympic Challenge: Wider, Deeper, Quicker, Better, Cheaper », *The American Statistician*, 61(1), p. 1 à 8.
- Kott, P.S. (2006), « Utilisation de la pondération par calage pour la correction de la non-réponse et des erreurs de couverture », *Techniques d'enquête*, 32(2), p. 149 à 160.
- Lee, S. et Valliant, R. (2009), « Estimation for volunteer panel Web surveys using propensity score adjustment and calibration adjustment », *Sociological Methods et Research*, 37(3), p. 319 à 343.
- Little, R.J.A. (2004). « To Model or Not Model? Competing Modes of Inference for Finite Population Sampling », *Journal of the American Statistical Association*, 99, 546 à 556.
- Loosveldt, G. et Sonck, N. (2008), « An evaluation of the weighting procedures for an online access panel survey », *Survey Research Methods*, 2(2), p. 93 à 105.
- Pfeffermann, D. et Sikov, A. (2011), « Imputation and Estimation under Nonignorable Nonresponse in Household Surveys with Missing Covariate Information », *Journal of Official Statistics*, 27(2), p. 181 à 209

- Rosenbaum, P.R. et Rubin, D.B. (1984), « Reducing bias in observational studies using subclassification on the propensity score », *Journal of the American Statistical Association*, 79, p. 516 à 524.
- Roshwalb, A., El-Dash, N. et Young, C. (2012), « Towards the Use of Bayesian Credibility Intervals in Online Survey Results », livre blanc produit par Ipsos Public Affairs et disponible à l'adresse http://www.ipsos-na.com/dl/pdf/knowledge-ideas/public-affairs/IpsosPA_POV_BayesianCredibilityIntervals.pdf.
- Royall, R.M. (1992), « Robustesse et optimalité de plan dans des modèles de prédiction pour populations finies », *Techniques d'enquête*, 18, 193 à 199.
- Särndal, C.-E., Swensson, B. et Wretman, J. (1992), *Model Assisted Survey Sampling*, New York: Springer-Verlag, Inc.
- Särndal, C.-E. et Lundquist, P. (2014), « Accuracy in Estimation with Nonresponse: A Function of Degree of Imbalance and Degree of Explanation », *Journal of Survey Statistics and Methodology*, 2, 361 à 387.
- Smith, T.M.F. (1976), « The Foundations of Survey Sampling: a Review ». *Journal of the Royal Statistical Society A*, 139, 183 à 195.
- Srivastav, A., Santibanez, T.A., Kahn, K.E., Zhai, Y., Greby, S.M. et coll. (2014), « National Early Season Flu Vaccination Coverage, United States, November 2014 ». Disponible à l'adresse <http://www.cdc.gov/flu/pdf/fluview/nifs-estimates-nov2014.pdf>.
- Statistique Canada (2002), « Cadre d'assurance de la qualité de Statistique Canada », disponible à l'adresse <http://www5.statcan.gc.ca/bsolc/olc-cel/olc-cel?catno=12-586-X et CHROPG=1 et lang=fra>.
- Statistique Canada (2009), « Statistique Canada : Lignes directrices concernant la qualité », disponible à l'adresse <http://www.statcan.gc.ca/pub/12-539-x/12-539-x2009001-fra.htm>.
- Thompson, S.K. (2002), *Sampling*, New York: John Wiley & Sons, Inc.
- Tourangeau, R., Conrad, F.G. et Couper M.P. (2013), *The Science of Web Surveys*, New York: Oxford University Press.
- Valliant, R. (2013), « Summary report of the AAPOR task force on non-probability sampling (Comment) », *Journal of Survey Statistics and Methodology*, 1, p. 105 à 111.
- Valliant, R. et Dever, J.A. (2011), « Estimating propensity adjustments for volunteer web surveys », *Sociological Methods et Research*, 40(1), p. 105 à 137.
- Valliant, R., Dever, J. A. et Kreuter, F. (2013), *Practical tools for designing and weighting sample surveys*, New York: Springer.
- Valliant, R. Dorfman, A. et Royall, R.M. (2000), *Finite Population Sampling and Inference: A Prediction Approach*, New York: John Wiley.
- Yeager, D.S., Krosnick, J.A., Chang, L., Javitz, H.S., Levendusky, M.S., Simpser, A. et Wang, R. (2011), « Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples », *Public Opinion Quarterly*, 75(4), p. 709 à 747. Disponible à l'adresse http://www.knowledgenetworks.com/insights/docs/Mode-04_2.pdf.
- Wang, W., Rothschild, D., Goel, S. et Gelman, A. (2015), « Forecasting Elections with Non-representative Polls », *International Journal of Forecasting*, à paraître.