

NOTE DE RECHERCHE/RESEARCH NOTE

Sondages probabilistes et autres créatures mythiques : usages et limites de la calibration des sondages non probabilistes

William Poirier¹ , Anne-Sophie Charest², Yannick Dufresne³, Alexandre Fortier-Chouinard³ et Nadjim Fréchet⁴

¹Département de science politique, Western University, 1151, Richmond St, London, ON, N6A 5C2, Canada, ²Département de mathématiques et de statistique, Université Laval, 1045, avenue de la Médecine, Québec, Québec, G1V 0A6, Canada, ³Département de science politique, Université Laval, 1030, avenue des Sciences-Humaines, Québec, Québec, G1V 0A6, Canada et ⁴Département de science politique, Université de Montréal, 3150, rue Jean-Brillant, Montréal, Québec, H3T 1N8, Canada

Auteur correspondant: William Poirier; Courriel: wpoirier@uwo.ca

(Reçu le 15 juillet 2024; révisé le 06 novembre 2024; accepté le 07 janvier 2025)

Résumé

Cette note de recherche vise à offrir une première introduction aux enjeux de la recherche par sondage, en particulier lorsqu'on utilise des données provenant de panels non probabilistes, comme les sondages en ligne. Nous expliquons le concept clé d'ignorabilité, qui aide à comprendre comment les biais de sélection peuvent affecter les résultats, et comment certaines techniques statistiques – comme la post-stratification et le *raking* – peuvent tenter de les corriger. À l'aide de simulations, nous montrons dans quels contextes ces méthodes peuvent fonctionner, et dans quels cas elles échouent. Les résultats suggèrent que les sondages non probabilistes présentent des limites importantes pour produire des estimations valides, mais qu'il existe aussi des pistes pour en améliorer l'usage, surtout dans le contexte actuel où ces données sont de plus en plus courantes en sciences sociales.

Abstract

This research note aims to provide an introductory overview of key issues in survey research, particularly when using data from non-probability panels such as online surveys. We introduce the central concept of ignorability, which helps clarify how selection bias can distort results and how certain statistical techniques – including post-stratification and raking – can be used to adjust for these biases. Through simulations, we explore the conditions under which these weighting methods succeed or fail. The results suggest that non-probability surveys face significant limitations in producing valid estimates but also highlight some promising ways to improve their use, especially given their growing prevalence in the social sciences.

Mots-clés: post-stratification; *raking*; ignorabilité; biais de sélection

Keywords: post-stratification; raking; ignorability; selection bias

Aujourd’hui, les spécialistes qui ont recours aux sondages sont confrontés à de nombreux obstacles : certaines personnes sont difficiles à joindre, les taux de réponse sont très faibles, et la qualité des réponses diminue. Ces problèmes, bien connus des spécialistes depuis un certain temps (Agans et al., 2021; Berinsky, 2017; de Heer et de Leeuw, 2002; Yeager et al., 2011), rendent la réalisation de sondages véritablement représentatifs extrêmement coûteuse (Hillygus, 2015). Dans ce contexte, de plus en plus d’entreprises et de chercheur·e·s se tournent vers des formes de sondages plus accessibles, comme les enquêtes en ligne, dont les participants ne sont pas sélectionnés selon un processus aléatoire (Couper, 2017; Olson et al., 2021; Kennedy et al., 2023). Ce changement dans la manière de constituer les échantillons soulève une question importante : Comment s’assurer que le phénomène observé ne soit pas dû à la sélection de nos cas, mais bien qu’il reflète une tendance se retrouvant dans la population étudiée?

Cette note de recherche propose un aperçu accessible, en français, d’une littérature majoritairement anglophone et très technique, qui s’est penchée sur cette question. Nous présenterons d’abord le principe d’ignorabilité, qui permet de mieux comprendre ce qu’est un biais de sélection. Nous explorerons ensuite comment certaines techniques de pondération peuvent corriger ce biais. Enfin, nous proposerons quelques pistes pour mieux utiliser des données qui sont parfois impossibles à corriger complètement. Pour illustrer ces enjeux, commençons par un exemple.

L’intuition de la prise d’échantillons

Disons que nous désirons estimer le taux de participation des prochaines élections fédérales. Demander à chaque électeur inscrit s’il compte aller voter, c’est-à-dire faire un recensement¹, est évidemment hors de portée. Cela est simplement trop coûteux et inutile. En effet, nous savons intuitivement qu’en sélectionnant soigneusement un certain nombre de répondants, nous serons en mesure d’estimer le taux de participation.² Or, la qualité de l’estimation à laquelle nous parviendrons découlera directement de la stratégie de sélection retenue, c’est-à-dire de notre échantillonnage.

Considérons un instant un échantillon de 1000 étudiant·e·s sur le campus d’une université.³ Immédiatement, nous pouvons émettre des doutes sur la représentativité de cet échantillon. Les étudiant·e·s universitaires ne représentent qu’un segment bien particulier des électeurs. Ils sont jeunes, éduqués et bien souvent plus engagés que le reste de la population. On peut donc s’attendre à ce que cet échantillon conduise à une surévaluation du taux de participation et qu’il y ait un biais dans l’estimation en raison de la sélection des cas, ce qu’on appelle un biais de sélection. Ce biais est dû à une erreur de couverture, soit le produit d’une liste de contacts ne comprenant pas tous les individus d’une population ou des individus en trop – nous avons trop de jeunes universitaires et il nous était impossible d’obtenir d’autres types d’électeurs (Weisberg, 2005).⁴

Plus formellement, on peut définir le biais de sélection ainsi : il s'agit de la différence entre la valeur réelle et la valeur estimée due à la sélection des cas. Si nous sélectionnons uniquement des individus ayant voté lors du précédent scrutin – ce qui est connu comme une sélection reposant sur la variable dépendante (Brady et Collier, 2010 : 349; Geddes, 1990) – nous ne serons pas surpris d'estimer des taux de participation très élevés. Formulé autrement, lorsque la règle d'échantillonnage n'assure pas la représentativité des cas, il est impossible pour les chercheur.e.s d'exclure la possibilité que la quantité mesurée (moyenne, écart-type, corrélation, etc.) soit due au processus de sélection.

La seule règle d'échantillonnage qui assure l'absence de biais de sélection est la sélection aléatoire des cas; on parle alors de sondage probabiliste,⁵ pourvu que la couverture soit appropriée et que le taux de non-réponse ne soit pas corrélé à la variable d'intérêt. Or, il est difficile d'obtenir une liste exhaustive d'individus de laquelle tirer aléatoirement des répondants sans être confronté à des problèmes de couverture. Même une liste de courriels exhaustive omettrait tout individu n'en ayant pas, ouvrant ainsi la possibilité à un biais de sélection. De plus, il y a toujours le problème de la non-réponse : les individus qui acceptent de répondre aux sondages diffèrent de ceux qui refusent de le faire, ce qui a une incidence sur l'inférence.⁶ La prochaine section porte sur les conditions à saisir pour obtenir un échantillon représentatif, ce qui nous amènera à réfléchir aux modalités à suivre afin d'identifier comment un sondage non probabiliste peut se montrer représentatif d'une population.

S'assurer de la représentativité des sondages

Quelles sont les conditions qui, une fois remplies, permettent aux chercheur.e.s d'ignorer le processus de sélection des cas comme source de biais potentiel? Le principe d'ignorabilité suggère deux conditions : (1) l'échangeabilité (*exchangeability*) et (2) la positivité (*positivity*).⁷ Reprenons notre estimation du taux de participation aux prochaines élections. Notre variable d'intérêt (ou variable dépendante) est la réponse à une question du type « Avez-vous l'intention de voter aux prochaines élections? » et peut être identifiée par la variable *Y*. Nous pouvons également dénoter l'âge des répondants avec la variable *X*, notre variable explicative (ou variable indépendante). Afin de caractériser la prise d'échantillon, imaginons une variable *Z* prenant la valeur de 1 pour les individus ayant été sélectionnés et 0 dans le cas contraire. *Z* possède ainsi une valeur pour chaque membre de la population.

La Figure 1 présente différents processus générateurs de données (la sélection des cas et la mesure de nos variables) pouvant intervenir lors de notre étude sur la participation électorale.⁸ Le panneau (a) de la Figure 1 représente une situation où les valeurs de *Z* ne sont pas influencées par une autre variable ayant une relation avec *Y*. C'est le résultat escompté lorsque nous utilisons un sondage probabiliste. Comme aucune autre influence n'a déterminé l'appartenance à l'échantillon, on pourrait théoriquement échanger des individus sélectionnés avec ceux exclus sans affecter significativement l'estimation du taux de participation. Ce scénario respecte alors la première condition, l'échangeabilité, soit la capacité d'échanger les unités

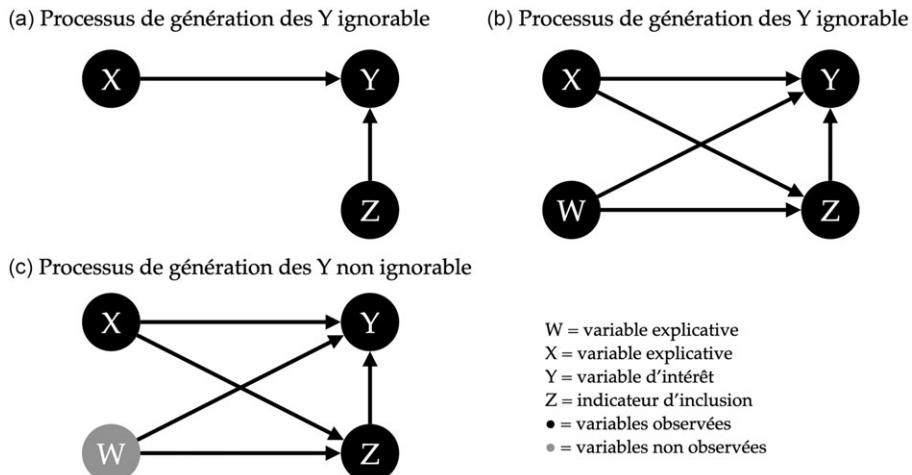


Figure 1. Représentation du mécanisme des biais de sélection.

échantillonnées avec celles ayant été exclues sans modifier de manière significative la quantité d'intérêt.

Si nous utilisons une règle non probabiliste pour la sélection des répondants, comme sélectionner aléatoirement⁹ des personnes se promenant sur le campus d'une université, nous pourrions nous retrouver dans une situation où l'échantillon surreprésente les 18 à 30 ans. Le panneau (b) illustre cette situation. Si nous tentions ici d'estimer le taux de participation, nous observerions un taux de participation plus élevé que dans la population générale. On pourrait alors parler d'un biais positif dans l'estimation. Toutefois, puisque Z est uniquement déterminé par l'âge (X) et que celui-ci est mesuré dans ce scénario, il serait possible d'estimer l'intention de vote sans biais pour chaque catégorie d'âge de X. On peut alors considérer que les individus inclus dans le sondage sont interchangeables avec les individus exclus à l'intérieur de chaque sous-groupe, ou qu'ils sont conditionnellement échangeables. Si, en plus, chaque sous-groupe de X est représenté dans l'échantillon – suivant notre exemple, des répondants de 18 à 30 ans, de 31 à 50 et de 51 ans et plus – nous allons satisfaire la condition de positivité. Celle-ci demande que toutes les valeurs possibles des variables exerçant une influence sur le processus de sélection (Z) ainsi que sur la variable dépendante (Y) soient présentes dans l'échantillon. Autrement dit, si nous respectons l'échangeabilité, mais que l'échantillon ne contient pas de personnes de plus de 30 ans, il devient impossible d'ajuster l'échantillon à l'aide de poids de pondération ou toute autre technique que ce soit. On ne peut pas inventer des individus qui sont absents du sondage. Ainsi, au panneau (b), puisque nous observons toutes les valeurs possibles de X et qu'il s'agit de la seule variable influençant Z et Y, le processus génératrice des données (PGD) est ignorable.

Reprendons le même sondage, mais imaginons cette fois que nous ignorons que l'échantillon a été collecté sur un campus. Nous constatons une surreprésentation de jeunes répondants puisque nous mesurons l'âge, mais nous ignorons d'où provient cette surreprésentation. Cette situation est présentée au panneau (c) de la Figure 1.

Tableau 1. Exemple de tableau de stratification

Université	Âge	Population (%)	Sondage (%)	Poids
Non	18–30	2,53	2,50	1,012
Non	31–50	25,14	24,50	1,026
Non	51+	40,80	32,70	1,248
Oui	18–30	17,14	22,80	0,752
Oui	31–50	6,89	9,30	0,741
Oui	51+	7,50	8,20	0,915

Dans le cas présent, en gris, W représente la variable non mesurée « aller à l'université » qui biaise à la hausse notre mesure du taux de participation. L'échangeabilité n'est donc plus uniquement conditionnelle à X comme au panneau (b), mais également conditionnelle à W . Puisque nous n'avons pas mesuré W , le processus générateur des données n'est plus ignorable et il nous est impossible de le corriger. Rappelons que même si nous avions mesuré W , il faudrait tenir compte de la condition additionnelle de positivité pour être en mesure de corriger notre estimation, c'est-à-dire avoir des non-universitaires dans l'échantillon. La prochaine section s'attarde à ce que l'on entend par « correction ».

Post-stratification et « raking »

Lorsque nous « corrigions » un sondage, nous voulons nous assurer que les proportions de certaines variables correspondent à celles retrouvées dans la population afin d'éviter de donner trop de poids (ou trop peu) à certaines catégories de la population et ainsi améliorer notre estimation. Les variables brisant la condition d'échangeabilité, mais pour lesquelles la positivité est respectée,¹⁰ revêtent ainsi un intérêt particulier. Différentes techniques permettent d'arriver à ce résultat. Nous nous concentrerons sur les deux techniques phares des méthodes de calibration¹¹ : la post-stratification et le « raking ».

La post-stratification¹² s'effectue en trois étapes : 1) la sélection des variables, 2) la création d'une table de stratification et 3) le calcul des poids de pondération. Comme nous l'avons vu, la sélection des variables est importante puisqu'il faut sélectionner les variables pour lesquelles la condition d'échangeabilité n'est pas respectée, sans quoi la pondération ne permettra pas de corriger le biais de sélection. Une condition pratique s'ajoute ici : il faut connaître la distribution de cette variable dans la population à l'étude. Ceci est nécessaire pour la deuxième étape, la création d'une table de stratification dont un exemple est présenté dans le Tableau 1.

Ce dernier présente les proportions croisées de l'âge et d'une scolarité universitaire dans la population et le sondage. Clairement, la règle d'échantillonnage n'a pas reproduit les mêmes ratios populationnels. Pour les corriger, il faut créer des poids de pondération qui, une fois multipliés aux unités du sondage, permettront de corriger l'estimation du taux de participation. Par exemple, le poids pour la catégorie diplôme universitaire de plus de 51 ans est calculé selon la formule suivante :

Tableau 2. Exemple de *raking* – Proportions du sondage

	18–30	31–50	51+	Total
Université Non	2,50	24,50	32,70	59,70
Université Oui	22,80	9,30	8,20	40,30
Total	25,30	33,80	40,90	100,00

$$w_i = \frac{\text{Population}(\%)_{ij}}{\text{Sondage}(\%)_{ij}} = \frac{7,50}{8,20} = 0,915 \quad (1)$$

où i représente un répondant et j la rangée du tableau de stratification correspondante. Remarquons que, si nous multiplions la proportion du sondage avec le poids de pondération correspondant, nous obtenons la proportion populationnelle ($0,915 \times 8,20 = 7,50$). Une moyenne concernant une variable d'intérêt (\bar{y}) pourra alors être corrigée en multipliant les valeurs de y par les poids de pondération :

$$\bar{y} = \frac{\sum_{i=1}^n y_i \times w_i}{n} \quad (2)$$

Malgré sa simplicité d'application, la post-stratification est rarement utilisée en pratique puisqu'elle exige que chaque catégorie croisée du tableau soit connue dans la population. Ces données croisées pouvant potentiellement être utilisées pour identifier les individus, elles sont rarement rendues accessibles par les entités gouvernementales les recueillant. Par exemple, Statistique Canada rend seulement disponibles les données agrégées, c'est-à-dire la proportion d'individus d'un certain âge, la proportion d'individus avec un diplôme universitaire, etc.

Le *raking* offre une solution alternative à la post-stratification, qui se distingue par l'utilisation de ces données agrégées. Cette technique emploie une méthode itérative; son algorithme est donc intégré à des progiciels (*packages*) comme *anesrake* (Pasek, 2018)¹³ développés en langage R. Si nous reprenons notre exemple, le Tableau 2 contient les proportions croisées des deux variables telles que recueillies dans le sondage. Nous savons à l'aide des données agrégées du recensement que la population contient 19,67 pour cent d'individus de 18 à 30 ans, 32,03 pour cent entre 31 et 50 ans, et 48,3 pour cent de 51 ans et plus, ainsi que 31,53 pour cent d'universitaires et 68,47 pour cent de non-universitaires. Pour retrouver ces quantités par *raking*, il faut d'abord multiplier les cellules de chaque rangée par leur ratio respectif du total à atteindre et du total actuel. Pour passer du Tableau 2 au Tableau 3, il faut multiplier chaque cellule de la rangée « Université Non » par $68,47 \div 59,70 = 1,146901$ et chaque cellule de la rangée « Université Oui » par $31,53 \div 40,30 = 0,782382$. De cette façon, on obtient le bon total pour ce qui est des proportions du statut d'universitaire. Or, les proportions pour l'âge sont toujours incorrectes.

Pour les corriger, il faut *ensuite* faire la multiplication des cellules par colonnes, et ce, à partir du Tableau 3. Le Tableau 4 présente les résultats de l'exercice qui a pour conséquence de déséquilibrer les ratios du statut d'universitaire. Il faudra donc répéter l'étape 1 puis l'étape 2, toujours à partir du dernier tableau produit, jusqu'à la

Tableau 3. Exemple de *raking* – Étape 1

	18–30	31–50	51+	Total
Université Non	2,87	28,10	37,50	68,47
Université Oui	17,84	7,28	6,42	31,53
Total	20,71	35,38	43,92	100,00

Tableau 4. Exemple de *raking* – Étape 2

	18–30	31–50	51+	Total
Université Non	2,72	25,44	41,24	69,41
Université Oui	16,95	6,59	7,06	30,59
Total	19,67	32,03	48,30	100,00

Tableau 5. Exemple de *raking* – Convergence

	18–30	31–50	51+	Total
Université Non	2,57	25,08	40,82	68,47
Université Oui	17,10	6,95	7,48	31,53
Total	19,67	32,03	48,30	100,00

Note: Quantités en pourcentage. Exemple inspiré de Lohr (2021 : 750–752).

convergence des deux types de ratios. Dans le cadre de cet exemple, il faudra répéter les deux étapes quatorze fois avant d'arriver à la convergence présentée au Tableau 5. Après avoir atteint la convergence, les poids de pondération seront calculés pour chaque cellule en divisant la quantité de convergence et la quantité originale. Ainsi, pour les universitaires de 18 à 30 ans, le poids de pondération sera $17,10 \div 22,80 = 0,750090$ afin de pallier leur surreprésentation dans l'échantillon.

Soulignons que le *raking* ne produit pas nécessairement les mêmes proportions croisées que l'on retrouverait dans une table de stratification. Cette approximation peut rendre le *raking* moins efficace que la post-stratification lorsque les croisements entre groupes ne sont pas répartis aléatoirement (Caughey et al., 2020). La section qui suit présente divers cas où la pondération se montre plus ou moins en mesure de corriger le biais induit.

Corriger le biais de sélection en pratique

Dans des conditions idéales, la pondération corrige parfaitement le biais de sélection. Une démonstration de ce point est réalisée dans l'Annexe A.1.1 à l'aide de simulations. Néanmoins, en pratique, notre capacité de correction est limitée par : (1) l'impossibilité de connaître le PGD; et (2) la disponibilité limitée des variables dont nous connaissons les valeurs dans la population. Typiquement, ces variables se limitent au profil sociodémographique des répondants – des variables à la racine de la plupart des phénomènes étudiés en sciences sociales. Or, celles-ci sont suffisantes

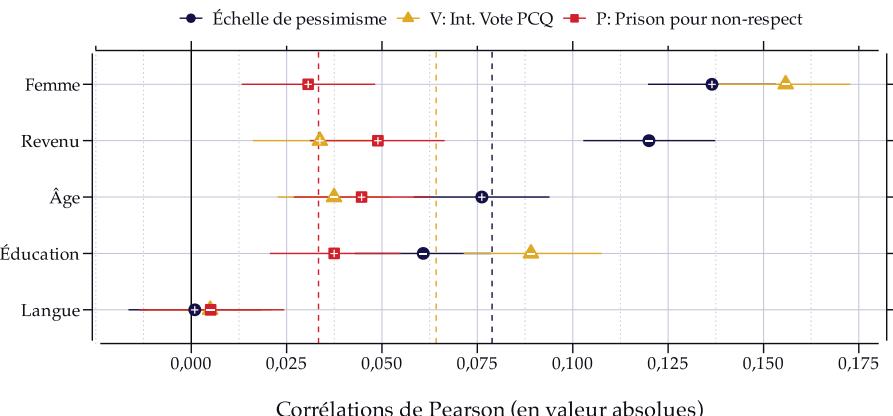


Figure 2. Force de la relation avec les variables d'intérêts.

Notes: Données issues de Projet Quorum et collectées du 26 janvier 2021 au 7 mars 2022; $n=12\,494$. Les signes à l'intérieur des points représentent la direction de la relation. Les lignes pointillées correspondent à la corrélation moyenne entre les variables de pondération et les variables d'intérêts. Intervalles de confiance (95 %) générées par bootstrap à l'aide du progiciel *confintr* (Mayer, 2023). Libellé des questions au Tableau A.2 dans l'Annexe A.

si et seulement si elles sont fortement corrélées à la variable dépendante (Y) et à la ou aux variables à la source du biais de sélection (Mercer *et al.*, 2017). Pour illustrer ce point, nous avons besoin de données répliquant plus fidèlement la réalité des interactions entre les variables présentes dans un sondage typique, variables pour lesquelles les valeurs populationnelles sont connues. Nous proposons l'utilisation des données d'un sondage à grand n qui nous servira de population d'étude et duquel nous pourrons prendre plusieurs échantillons. Les données de *Projet Quorum*¹⁴ comprennent les réponses complètes de 12 494 répondants collectées du 26 janvier 2021 au 7 mars 2022 sur certains enjeux phares du contexte de la pandémie de COVID-19 – les applications de traçage, les restrictions sanitaires et les valeurs démocratiques. Considérons donc ce sondage comme la population à l'étude et imaginons vouloir estimer le niveau de pessimisme (Y) dans cette même population (une échelle de 0 à 1).

Imaginons également que nous avons accès à 3 sondages non probabilistes de 1000 individus et à la distribution dans la population (1) du genre, (2) du revenu, (3) de l'âge, (4) de l'éducation et (5) de la langue. Le premier sondage, nous le voyons dans les données, surreprésente les femmes. Nous devrions ainsi être en mesure de corriger ce biais si nous respectons la positivité – plus précisément si nous avons des hommes dans l'échantillon. Or, les deux autres surreprésentent respectivement les individus ayant l'intention de voter pour le Parti conservateur du Québec (PCQ) et les individus appuyant l'idée que l'État devrait emprisonner les individus ne respectant pas les mesures sanitaires. La Figure 2 présente les relations entre nos variables de pondération, l'échelle de pessimisme et les variables causant le biais de sélection dans la population.

Notons d'abord que parmi les variables sociodémographiques, seuls le genre (F) et le revenu possèdent une corrélation importante avec l'échelle de pessimisme. Ceci indique qu'il nous sera impossible de corriger significativement un biais sur l'échelle de pessimisme si nous nous restreignons à l'éducation et à la langue, par exemple.

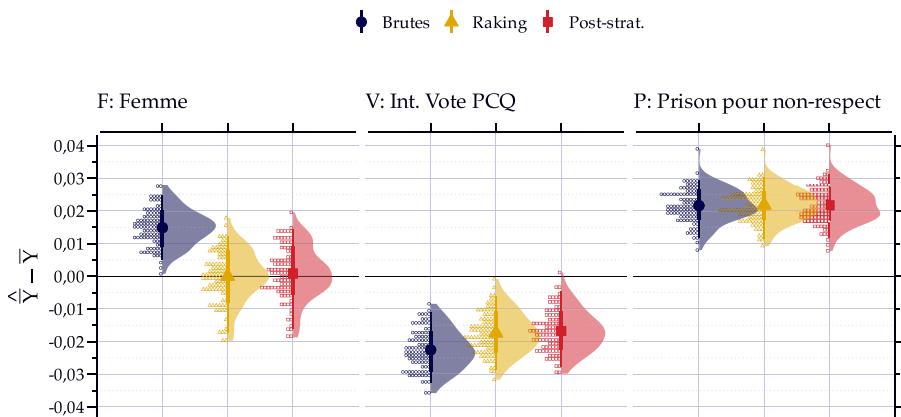


Figure 3. Effet des techniques de correction sur l'estimation du niveau de pessimisme.

Note: Données simulées à partir de Projet Quorum. 100 échantillons de 1000 unités par cas. $\alpha = -30$, $\beta = 1,7$. Variables : Y l'échelle de pessimisme, F s'identifier comme femme, V intention de voter pour le PCQ, P opinion sur l'emprisonnement des individus ne respectant pas les mesures sanitaires. Corrélations : $\text{Corr}(F, Y) = 0,14$; $\text{Corr}(V, Y) = -0,17$; $\text{Corr}(P, Y) = 0,25$.

Le premier sondage surreprésente les femmes. Puisque le genre est la variable sociodémographique la plus corrélée avec le pessimisme dans ce contexte, nous pouvons nous attendre à ce que l'estimation soit biaisée à la hausse. Or, puisqu'il est inclus dans les variables de pondération, nous devrions être en mesure de corriger ce biais.

Notons ensuite que ces mêmes variables sont moins corrélées en moyenne avec l'intention de vote pour le PCQ (V) et encore moins corrélées en moyenne avec l'emprisonnement lié au non-respect des mesures sanitaires (P). Ceci implique que si ces variables (V et P) sont fortement corrélées avec le pessimisme – et elles le sont ($\text{Corr}(V, Y) = -0,17$; $\text{Corr}(P, Y) = 0,25$) – nous ne pourrons que partiellement réduire le biais à l'aide des variables sociodémographiques. Afin d'estimer le pouvoir de correction de ces variables sociodémographiques dans le contexte des trois sondages, simulons la prise de 100 échantillons de 1000 personnes pour chaque contexte (biais causé par F , V et P).¹⁵ La Figure 3 présente la distribution des biais induits, leur moyenne, ainsi que ce qu'il en advient lorsqu'il est corrigé par raking et post-stratification.

Comme prévu, il est possible de constater un biais dans l'estimation du pessimisme dans chacun des cas, positif lorsque la corrélation avec Y est positive et négatif dans le cas contraire. Pour le premier sondage, la pondération basée sur les variables sociodémographiques suffit à corriger le biais en moyenne.¹⁶ Ce n'est pas surprenant, puisque le biais vient du genre et que nous corrigeons pour cette variable. Lorsque la variable causant le biais ne se retrouve pas dans les variables de correction, comme dans les deux autres cas, le pouvoir de correction de la pondération devient une fonction de la relation entre les variables de correction et la variable causant le biais. On constate que nous arrivons à corriger l'estimation partiellement dans le deuxième sondage et pas du tout dans le troisième.

La pondération a donc la capacité de corriger parfaitement les biais de sélection, même lorsque le processus d'échantillonnage n'est pas aléatoire. Or, il faut connaître

le PGD afin d'être en mesure d'identifier les variables qui empêchent de respecter la condition d'échangeabilité. Ceci est une quasi-impossibilité pour les sondages non probabilistes. En effet, sans contrôle du PGD et avec des théories limitées sur ce à quoi il peut ressembler, les chercheur.e.s ne peuvent présenter qu'une présomption théorique limitée par les variables disponibles dans les données du recensement. Même lorsque le PGD hypothétique est proche de la réalité, les chercheur.e.s sont limité.e.s par les données du recensement. Les modèles de prédiction des attitudes allaient au-delà des variables sociodémographiques dès 1960 (Campbell et al., 1960), alors l'idée qu'un seul vecteur de pondération basé sur des variables sociodémographiques suffise à corriger même une seule estimation apparaît absurde.

Cet état de fait est reflété par des études appliquées s'étant penchées sur la qualité des données non probabilistes comme celles de Dutwin et Buskirk (2017), Felderer et al. (2019), MacInnis et al. (2018), Pasek et Krosnick (2010), Stern et al. (2017), et Yeager et al. (2011). Toutes ces études comparent des sondages probabilistes à des sondages non probabilistes et toutes concluent que la pondération n'arrive pas à effacer les biais induits. Seuls quelques cas rapportent un tel succès : Ansolabehere et Schaffner (2014), Gelman et al. (2016), et Wang et al. (2015).¹⁷ Ces études témoignent bien plus de la difficulté de la tâche que d'une solution miracle. Encore une fois, l'incapacité à effacer le biais de sélection n'est pas due à une limite technique ou statistique de la correction, mais bien à un échec de la sélection des variables de pondération.

Un plus grand n , ça vaut la peine?

Pour un échantillon probabiliste, augmenter la taille (n) permet de réduire l'erreur d'échantillonnage pour la moyenne d'une variable \bar{Y} selon $Z \times \frac{\sigma_Y}{\sqrt{n}}$ (Lohr, 2021). Cette même intuition ne doit toutefois pas être transférée au contexte non probabiliste. À la lumière du principe d'ignorabilité, il est effectivement possible de définir le biais de sélection comme la différence entre la moyenne estimée (\hat{Y}) et la moyenne de la population (\bar{Y}) en fonction de la taille de l'échantillon (n), la taille de la population (N), l'écart-type de Y dans la population (S_Y), ainsi que la corrélation entre Z et Y (Lohr, 2021).¹⁸ Formellement, l'équation se comprend comme suit :

$$\text{Biais}(\hat{Y}) = E_Z[\hat{Y} - \bar{Y}] = E_Z[\text{Corr}(Z, Y)] \times \sqrt{\frac{N-1}{n} \left(1 - \frac{n}{N}\right)} \times S_Y \quad (3)$$

Si l'équation peut paraître intimidante, il s'agit en fait de la multiplication de trois termes : 1) S_Y est l'écart-type de la variable d'intérêt dans la population; 2) le terme sous la racine carrée est essentiellement un ratio entre la taille de la population et la taille de l'échantillon; et 3) $\text{Corr}(Z, Y)$ correspond à la corrélation entre la variable d'intérêt et le processus de sélection. Le pouvoir de l'aléatoire est alors évident : si Z est déterminée aléatoirement, alors sa corrélation avec Y sera 0 et le biais disparaîtra.

Bien que les simulations présentées plus haut le permettent, il est rare d'avoir le contrôle sur l'écart-type de Y ou sur sa corrélation avec Z (dans le contexte d'un échantillon non probabiliste). Ces éléments sont plutôt fixés par la nature lors de la prise de l'échantillon. Dans ce contexte, l'unique valeur exerçant une influence sur le biais de \hat{Y} et qui peut être contrôlée est n . À première vue, le terme central

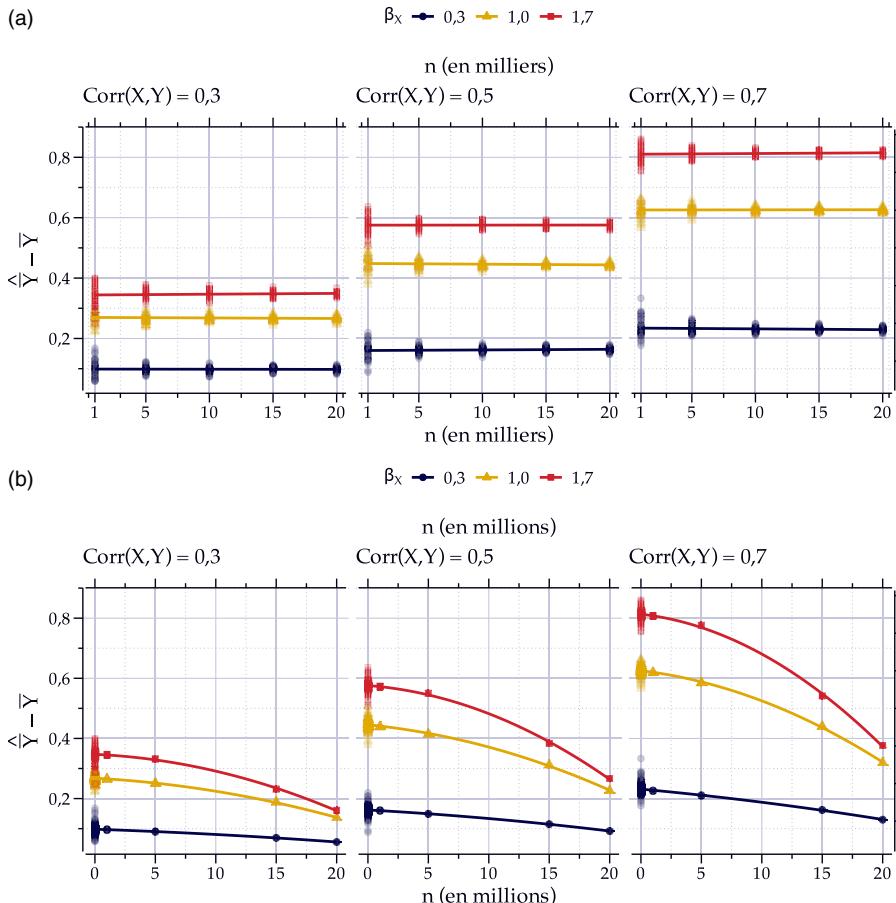


Figure 4. Effet de la taille d'échantillon sur la magnitude de $\widehat{Y} - \bar{Y}$.

Note: Données issues d'une simulation. Tendances générées par régression linéaire (OLS). Points correspondant à la moyenne du biais brut des itérations. Dans tous les cas, $N = 31\,000\,000$, Y est continu et distribué normalement ($\bar{Y} = 0, \sigma_y = 1$), X est une échelle de Likert à 4 niveaux (0, 1, 2, 3), et Z est binaire (0, 1), indiquant l'inclusion dans l'échantillon.

$\sqrt{\frac{N-1}{n}(1-\frac{n}{N})}$ diminue lorsque n augmente, ce qui justifierait l'intuition selon laquelle augmenter n réduit $\widehat{Y} - \bar{Y}$. Simulons une prise d'échantillon auprès de 31 000 000 d'unités dont le PGD correspond à celui du panneau (b) de la Figure 1 (X influence Y et Z , et Z influence Y) en variant le n (de 1 000 à 20 000 000) pour tester cette intuition.

Le panneau (a) de la Figure 4 est catégorique : le seul effet positif d'une augmentation du n d'un échantillon non probabiliste est une diminution de la variance du biais et non pas une diminution du biais lui-même. En effet, indépendamment de la Corr(X, Y) ou de la force de la relation entre X et Z (β_X), le biais moyen en espérance demeure le même alors que n augmente. Il ne s'agit

d'ailleurs pas de petites augmentations de n ; même à 20 000 unités, le biais ne bouge pas. Pour percevoir une diminution du biais, il faut un échantillon contenant 5 000 000 d'unités, comme le montre le panneau (b). Or, un n de cette magnitude n'est pas suffisant pour corriger complètement le biais. En fait, la différence est négligeable dans les cas où le biais d'origine est élevé. Pire encore, à $n = 15\,000\,000$, soit approximativement la moitié de la population simulée, le biais n'est toujours pas corrigé. Augmenter le n n'est donc pas une stratégie viable à la correction des sondages non probabilistes. Ceci est dû au fait que $\text{Corr}(Z, Y)$ est partiellement dépendante de n , de sorte que la corrélation augmente alors que n augmente.¹⁹ Augmenter n n'a donc de mérite que lorsque l'objectif est de respecter la condition de positivité, dans le but, par exemple, de s'assurer d'avoir suffisamment de francophones dans un échantillon canadien.

Que doit-on faire alors?

En l'absence de connaissance du PGD, la sélection aléatoire constitue la meilleure solution pour éviter le biais de sélection. Il faut toutefois s'assurer d'avoir une couverture appropriée et que la non-réponse ne soit pas corrélée à la variable d'intérêt. Or, l'accès à des échantillons probabilistes est souvent impossible. Il existe néanmoins une avenue permettant de mobiliser les données issues d'échantillons non probabilistes : les analyses de sensibilité.²⁰

Développées par Rosenbaum (2005), celles-ci ont pour objectif de donner une idée de la vulnérabilité face à un éventuel biais de variable omise des résultats d'une analyse causale basée sur des données observationnelles. En évaluant la force qu'une variable non mesurée devrait avoir pour nullifier l'effet mesuré, voire l'inverser, et en la comparant avec la corrélation d'une des variables mesurées avec la variable d'intérêt, les analyses de sensibilité permettent de relativiser l'influence d'un potentiel biais de variable omise (Cinelli et Hazlett, 2020). Ce faisant, il est possible de défendre les résultats d'une analyse causale réalisée à partir de données observationnelles devant la critique du biais de variable omise. Certes, la relation n'est pas identifiée, mais l'analyse peut permettre de découvrir qu'une variable omise devrait avoir une relation beaucoup trop grande avec Y pour significativement influencer la mesure de l'effet.

Hartman et Huang (2024) adaptent l'analyse de sensibilité au contexte descriptif en se concentrant sur la différence entre les poids de pondération estimés et les poids de pondération idéaux ($\varepsilon = w - w^*$). L'équation 1 est alors reformulée ainsi :

$$\text{Biais}(\bar{y}) = E[\bar{y} - \bar{Y}] = E[\text{Corr}(\varepsilon, Y)] \times \sqrt{S_Y \times S_w \times \frac{R_\varepsilon^2}{1 - R_\varepsilon^2}} \quad \forall R_\varepsilon^2 < 1 \quad (4)$$

où R_ε^2 est le ratio de variation en w^* expliqué par ε et est < 1 . L'idée est de faire varier $\text{Corr}(\varepsilon, Y)$ afin d'identifier les conditions sous lesquelles l'ajustement des poids de pondération est insuffisant. Ceci permet alors d'identifier les situations où pondérer en utilisant les variables disponibles du recensement suffit à corriger le biais et les situations où des analyses descriptives sont à éviter. Or, cette approche repose sur les chercheur.e.s pour poser les valeurs de $\text{Corr}(\varepsilon, Y)$ et de R_ε^2 , ce qui ne

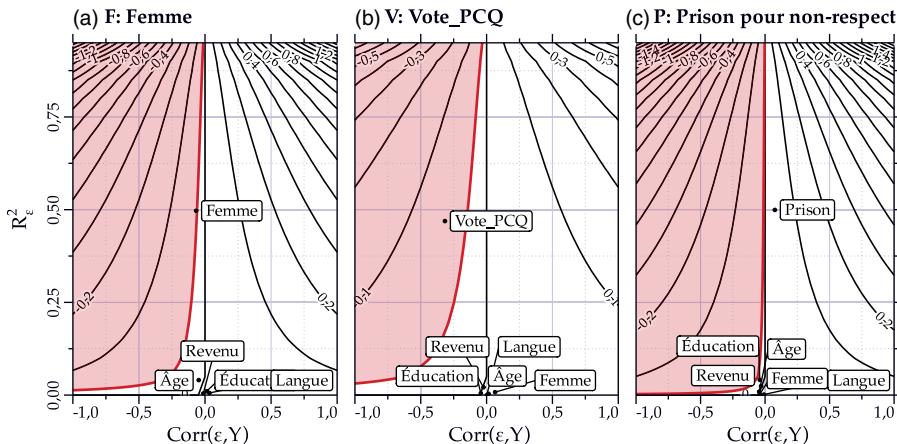


Figure 5. Sensibilité de l'estimation à un biais de variable omise.

Note: Données simulées à partir de Projet Quorum. Analyse réalisée sur l'un des échantillons de la Figure 3 pour chaque variable utilisée pour biaiser l'échantillon. Poids de pondérations produit à partir de *raking* sur les variables socio-démographiques utilisées plus haut. R^2_ε est le ratio de variation en w^* expliqué par ε .

régle pas totalement le problème, car ces décisions reposent toujours sur le contexte théorique et sur l'idée que les chercheur.e.s se font du PGD.

Pour diminuer les effets découlant de ce problème, Hartman et Huang (2024) produisent des graphes de contours (*contour plots*). Ceux-ci permettent de visualiser plusieurs combinaisons de $\text{Corr}(\varepsilon, Y)$ et de R^2_ε ainsi que leur impact sur le biais théorique ($E[\bar{y} - \bar{Y}]$). Prenons comme exemple l'un des échantillons biaisés sur le genre ayant été produit à la Figure 3. Suivant la procédure de Hartman et Huang (2024), nous produisons des poids de pondération à l'aide de *raking*. Rappelons que, dans ce scénario, nous avons uniquement accès aux variables sociodémographiques pour produire nos poids de pondération. Nous obtenons alors un échantillon produisant une moyenne de pessimisme ($\widehat{\bar{Y}}$) est à 0,39 et la moyenne pondérée ($\widehat{\bar{Y}}^*$) est à 0,42 là où nous savons que la valeur populationnelle est à 0,415. Nous pouvons également calculer la variance de notre estimation (S_Y) ainsi que la variance des poids de pondération (S_w).

Puisque $\text{Corr}(\varepsilon, Y)$ et de R^2_ε varient respectivement entre -1 et 1 et 0 et 1, nous sommes en mesure de produire un biais théorique pour chaque combinaison possible. Ceux-ci sont présentés à la Figure 5 panneau (a) et correspondent aux lignes noires à l'intérieur du graphique. Ainsi, un biais de +0,2 peut être produit par toutes les combinaisons de valeurs de $\text{Corr}(\varepsilon, Y)$ et de R^2_ε se trouvant sur la ligne. Seul, ce graphique n'est pas très informatif. Le véritable tour de force de Hartman et Huang (2024) est d'y superposer deux informations additionnelles.

D'abord, une zone critique établie par le chercheur en fonction du contexte spécifique au sondage. Imaginons que notre sondage est commandé par le gouvernement afin d'évaluer s'il est nécessaire de réaliser une intervention auprès de la population et que cette intervention sera faite si le niveau de pessimisme atteint

0,4 en moyenne. Un biais de $0,39 - 0,4 = -0,01$ et plus viendrait alors changer la prise de décision; il s'agit de la zone représentée en rouge sur la figure.

Ensuite, il est possible de calculer la variance expliquée des poids de pondération (R^2_ε) par le retrait successif de chaque variable de pondération. Ainsi, à tour de rôle, les poids de pondération sont recalculés sans utiliser la variable *Femme*, puis *Éducation*, puis *Revenu*, etc. De cette façon, il est possible d'évaluer la sensibilité de l'estimation à l'inclusion de différentes variables dans le calcul des poids de pondération. Au panneau (a) de la Figure 5, il est clair que l'estimation est beaucoup plus sensible à l'inclusion de la variable *Femme*. En ce sens, la variance expliquée (R^2_ε) par cette variable est beaucoup plus élevée que celle expliquée par les autres, ce qui fait en sorte que, lorsque *Femme* est exclue des variables de pondération, l'estimation se retrouve à l'intérieur de la zone critique. Cela est évident dans le cas présent, puisque l'échantillon surreprésente les femmes. Or, pour un véritable sondage, cette technique permet d'identifier le niveau de sensibilité de nos estimations aux choix de variables de pondération.

Si des estimations populationnelles de sondage de qualité (comme ceux de l'Étude électorale canadienne) sont disponibles, il est également possible de réaliser cet exercice pour des variables ne se retrouvant pas dans les données du recensement. C'est ce que nous faisons aux panneaux (b) et (c) de la Figure 5. Nous simulons d'abord un sondage de qualité en prenant un échantillon aléatoire des données de *Projet Quorum*. Ceci nous permet d'estimer les proportions populationnelles de *Vote_PCQ* et *Prison*, respectivement. Puis nous répétons l'exercice réalisé plus haut pour *Femme*.

Au terme de l'exercice, nous concluons que l'estimation sous un biais provenant de *Vote_PCQ* est très sensible à son exclusion alors que le biais provenant de *Prison* est dans la « bonne » direction, c'est-à-dire que le biais causé augmente le niveau de pessimisme, éloignant l'estimation de la zone critique.

Conclusion

Nous nous sommes intéressés à la question de la correction des sondages non probabilistes à l'aide de la post-stratification et du *raking*. En raison du biais de sélection, leur utilisation est problématique lorsqu'il est question de quantités descriptives. Nous avons établi le principe d'ignorabilité comme l'appareillage théorique devant guider l'échantillonnage. Ce dernier précise que, pour être ignorable, un PGD doit respecter deux conditions : l'échangeabilité et la positivité. La condition d'échangeabilité est satisfaite lorsqu'il est possible d'échanger n'importe quelles valeurs de $Y_{Z=1}$ pour n'importe quelles valeurs de $Y_{Z=0}$ sans substantiellement influencer la moyenne estimée \bar{Y} en espérance. La positivité est quant à elle respectée lorsque l'entièreté des valeurs possibles des variables et combinaisons de variables entrant dans le PGD est mesurée.

Ces conditions à la correction ont été testées empiriquement. Lorsqu'elles sont observées, les techniques de correction nullifient l'effet du biais de sélection. Ceci pointe vers les limites pratiques et non techniques de la pondération. En effet, être en mesure de définir parfaitement le PGD est, par définition, impossible. À cela s'ajoute le fait que certaines variables ne sont tout simplement pas accessibles en raison d'un manque de ressources ou d'enjeux éthiques; de plus, au-delà des

variables sociodémographiques contenues dans le recensement, les quantités populationnelles sont le plus souvent inconnues.

Que donc faut-il retenir au terme de cet exercice? D'abord, la réalité de la recherche effectuée en sciences sociales empêchera la correction de la plupart des inférences descriptives issues d'échantillons non probabilistes. Ceci demeurera le cas aussi longtemps que les variables pour lesquelles les valeurs populationnelles sont connues ne seront pas étendues au-delà des variables sociodémographiques. Néanmoins, et il s'agit du deuxième point clé, cela n'implique pas le rejet dogmatique de l'utilisation des données non probabilistes. Celles-ci peuvent se montrer fort utiles lorsque l'on désire tester le libellé de certaines questions avant leur utilisation, réaliser une expérience de sondage ou lorsque l'ignorabilité est respectée. Cela n'est pas sans rappeler les sages paroles de King et al. (1994: 27) : « Social scientists often find themselves with problematic data and little chance to acquire anything better; thus, they have to make the best of what they have ». En ce sens, relativiser le biais attendu à l'aide d'analyses de sensibilité est peut-être la meilleure façon de valoriser ce type de données.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0008423925100528>

Conflit d'intérêt. Les auteurs n'ont aucun conflit d'intérêt à déclarer.

Notes

1 Le Tableau A.3 dans l'Annexe A recense tous les concepts abordés dans cette note ainsi que leur définition.

2 Ce résultat statistique aujourd'hui devenu intuition fut vigoureusement débattu au début du 20^e siècle et ne s'est établi dans les pratiques commerciales que vers la fin des années 1940 (Brewer, 2013; Converse, 1987; Kjaer, 1897).

3 Nous utilisons ici l'exemple d'un échantillon recruté directement par les chercheur·e·s, mais dans la pratique, ceux-ci font souvent appel à des panels web gérés par des entreprises privées. Le principe d'ignorabilité s'applique aussi : si le panel est non probabiliste, l'échantillon sera biaisé, même s'il est tiré au hasard à l'intérieur du panel.

4 Nous aurions pu obtenir un résultat similaire à l'aide d'un échantillon aléatoire où seuls de jeunes universitaires auraient accepté de répondre. Le biais de sélection aurait alors été causé par une *erreur de non-réponse* (Weisberg, 2005).

5 Ceci peut prendre la forme d'une sélection aléatoire simple sans remplacement où on tire à la pique n individus d'une liste exhaustive ou d'autres méthodes plus complexes. L'important est que la probabilité de sélection de chaque unité de la population à l'étude soit connue et que cette probabilité soit non nulle pour tous (Henry, 1990; Lohr, 2021).

6 L'inférence est le processus d'utilisation des faits que nous connaissons pour en apprendre davantage sur les faits que nous ne connaissons pas, ou notre habileté à généraliser à une population des faits basés sur un échantillon (King, Keohane, et Verba, 1994).

7 D'abord développé pour le traitement des données manquantes par Rubin (1978), le principe d'ignorabilité est une approche qui se concentre sur le processus générateur de données (*data generating process*, PGD) et permet de définir le biais de sélection ainsi que ses causes (Little, 1993; Mercer et al., 2017).

8 Pour une discussion complète de comment lire et interpréter les graphes orientés acycliques comme celui de la Figure 1, voir Elwert (2013) ou le chapitre 6 de Arel-Bundock (2020).

9 Rappelons ici que la sélection aléatoire garantit l'ignorabilité du PGD uniquement dans deux cas : (1) si le taux de couverture et le taux de réponse sont à 100 pour cent *ou* (2) si la couverture est appropriée et la non-réponse n'est pas corrélée à la variable dépendante.

10 C'est du moins le cas dans le contexte d'un sondage non probabiliste. En effet, la calibration a été développée pour corriger l'*erreur d'échantillonnage* due au processus aléatoire. Celle-ci est une fonction de l'écart-type de la variable d'intérêt dans la population ainsi que de la taille de l'échantillon, l'intuition étant que plus l'échantillon est grand, plus l'erreur d'échantillonnage diminue (Lohr, 2021). C'est ce qui explique que, parfois, la moyenne d'un échantillon aléatoire dévie de la moyenne de la population. Puisque la probabilité d'être sélectionné est connue pour chaque membre de la population, il est possible de corriger cette erreur en pondérant les résultats à l'aide de l'inverse de cette même probabilité. Dans le contexte d'un sondage non probabiliste, la calibration est plutôt utilisée pour corriger les erreurs de couverture et de non-réponse.

11 D'autres techniques d'ajustement comme les scores de propension, le *matching* et la régression multiniveau avec post-stratification (*multi-level regression with post-stratification* ou *MRP*) ont été développées afin de répondre aux mêmes besoins d'ajustement que les outils de calibration (Mercer *et al.*, 2017). Voir Valliant et Dever (2011) pour une discussion des scores de propension, Ansolabehere et Schaffner (2014) pour une application réussie du *matching*, et Wang *et al.* (2015) comme exemple phare de la MRP.

12 Pour une discussion plus technique de la post-stratification, voir Holt et Smith (1979), Smith (1991) et Little (1993). Cette dernière source offre une perspective bayésienne.

13 Pour une revue détaillée du fonctionnement d'*anesrake*, voir DeBell et Krosnick (2009).

14 Le libellé des questions de Projet Quorum se trouve au Tableau A.2 dans l'Annexe A.

15 Voir la section A.1.3 pour plus d'information quant à la réalisation des simulations d'échantillons non probabilistes.

16 C'est-à-dire que dans la plupart des cas, on peut corriger en très grande partie l'estimation, mais que ce n'est pas garanti à chaque fois. Remarquons à cet effet les quelques échantillons de la Figure 3 qui, malgré la pondération, se retrouvent à surestimer ou sous-estimer de façon substantielle le niveau de pessimisme.

17 Leur succès ne peut que partiellement être attribué à la sophistication de leurs méthodes. Celles-ci en doivent beaucoup au fait que les variables de pondération sont fortement reliées aux variables d'intérêts et à leur accès à des échantillons à très grand n , 35 000 000 d'unités dans le cas de Wang *et al.* (2015). Ceci leur permet d'élargir leur table de stratification tout en maintenant plusieurs répondants dans chaque cellule (augmenter le n pour respecter la positivité).

18 L'estimateur retrouvé dans Lohr (2021) est issu et adapté de Meng (2018), qui le présente formellement et empiriquement en étudiant le cas de l'élection présidentielle américaine de 2016.

19 Ceci pour un échantillon fini, une population finie et pour des corrélations de Pearson. Plus de détails à la section A.1.2 de l'Annexe A.

20 Un-e évaluateur-trice a proposé d'inclure systématiquement le vote dans les pondérations. Bien que cette variable soit centrale en science politique et qu'on puisse en principe estimer sa distribution, nous ne faisons pas cette recommandation pour trois raisons : (1) la stabilité du vote au Canada est incertaine (Gidengil, 2022), rendant toute estimation rapidement obsolète après une élection ; (2) les résultats électoraux ne reflètent que les votants, or il est difficile d'identifier avec précision qui vote dans un sondage ; (3) l'intention de vote est souvent mal mesurée, ce qui limite son utilité (Green *et al.*, 2002).

Références

- Agans, Robert P., Donglin Zeng, Bonnie E. Shook-Sa, Marcella H. Boynton, Noel T. Brewer, Erin L. Sutfin, Adam O. Goldstein, *et al.* 2021. Using Social Networks to Supplement RDD Telephone Surveys to Oversample Hard-to-Reach Populations: A New RDD + RDS Approach. *Sociological Methodology* 51 (2): 270–89.
- Ansolabehere, Stephen, et Brian F. Schaffner. 2014. Does Survey Mode Still Matter? Findings from a 2010 Multi-Mode Comparison. *Political Analysis* 22 (3): 285–303.
- Arel-Bundock, Vincent. 2020. *Analyse causale et méthodes quantitatives: une introduction avec R, Stata et SPSS*. Presses de l'Université de Montréal.
- Berinsky, Adam J. 2017. Measuring Public Opinion with Surveys. *Annual Review of Political Science* 20: 309–29.
- Brady, Henry E., et David Collier. 2010. *Rethinking social inquiry: Diverse tools, shared standards*. Rowman & Littlefield Publishers.

- Brewer, Ken. 2013. Three Controversies in the History of Survey Sampling. *Survey Methodology* 39 (2): 249–63.
- Campbell, Angus, Philip E. Converse, Warren E. Miller, et Donald E. Stokes. 1960. *The American Voter*. University of Chicago Press.
- Caughey, Devin, Adam J. Berinsky, Sara Chatfield, Erin Hartman, Eric Schickler, et Jasjeet S. Sekhon. 2020. Target Estimation and Adjustment Weighting for Survey Nonresponse and Sampling Bias. 1^{re} éd. Cambridge University Press.
- Cinelli, Carlos, et Chad Hazlett. 2020. Making Sense of Sensitivity: Extending Omitted Variable Bias. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 82 (1): 39–67.
- Converse, Jean M. 1987. *Survey Research in the United States : Roots and Emergence*. Berkeley: University of California Press.
- Couper, Mick P. 2017. New Developments in Survey Data Collection. *Annual Review of Sociology* 43: 121–45.
- De Heer, Wim, et Edith D. de Leeuw. 2002. Trends in household survey nonresponse: A longitudinal and international comparison. In *Dans Survey Nonresponse*, édité par Robert M. Groves, Don A. Dillman, John L. Eltinge, et Roderick J. A. Little, 41–54. Wiley.
- DeBell, Matthew, et Jon A. Krosnick. 2009. Computing weights for American national election study survey data. nes012427. Ann Arbor; Palo Alto: ANES Technical Report Series.
- Dutwin, David, et Trent D. Buskirk. 2017. Apples to Oranges or Gala Versus Golden Delicious? Comparing Data Quality of Nonprobability Internet Samples to Low Response Rate Probability Samples. *Public Opinion Quarterly* 81 (S1): 213–39.
- Elwert, Felix. 2013. Graphical causal models. Dans *Handbook of Causal Analysis for Social Research*, édité par Stephen L. Morgan, Springer. 245–73.
- Felderer, Barbara, Antje Kirchner, et Frauke Kreuter. 2019. The Effect of Survey Mode on Data Quality: Disentangling Nonresponse and Measurement Error Bias. *Journal of Official Statistics* 35 (1): 93–115.
- Geddes, Barbara. 1990. How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics. *Political Analysis* 2: 131–50.
- Gelman, Andrew, Sharad Goel, Douglas Rivers, et David Rothschild. 2016. The Mythical Swing Voter. *Quarterly Journal of Political Science* 11 (1): 103–30.
- Gidengil, Elisabeth. 2022. Voting Behaviour in Canada: The State of the Discipline. *Canadian Journal of Political Science* 55 (4): 916–38.
- Green, Donald P., Bradley Palmquist, et Eric Schickler. 2002. *Partisan Hearts and Minds: Political Parties and the Social Identities of Voters*. New Haven; London: Yale University Press.
- Hartman, Erin, et Melody Huang. 2024. Sensitivity Analysis for Survey Weights. *Political Analysis* 32 (1): 1–16.
- Henry, Gary T. 1990. *Practical Sampling*. Sage Publications.
- Hillygus, D. Sunshine. 2015. The Practice of Survey Research: Changes and Challenges. Dans *New directions in public opinion*, édité par Adam Berinski, Routledge. 21–40.
- Holt, David, et T. M. Fred Smith. 1979. Post-stratification. *Journal of the Royal Statistical Society Series A: Statistics in Society* 142 (1): 33–46.
- Kjaer, Anders N. 1897. The representative method of statistical surveys. Papers from the Norwegian Academy of Science and Letters, II. The Historical, philosophical Section.
- King, Gary, Robert O. Keohane, et Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton University Press.
- Little, Roderick J. A. 1993. Post-stratification: a modeler's perspective. *Journal of the American Statistical Association* 88 (423): 1001–12.
- Lohr, Sharon L. 2021. *Sampling: Design and Analysis*. New York: CRC press.
- MacInnis, Bo, Jon A. Krosnick, Annabell S. Ho, et Mu-Jung Cho. 2018. The accuracy of measurements with probability and nonprobability survey samples: Replication and extension. *Public Opinion Quarterly* 82 (4): 707–44.
- Mayer, Michael. 2023. confintr: Confidence Intervals. <<https://mayer79.r-universe.dev/confintr>>.
- Meng, Xiao-Li. 2018. Statistical Paradises and Paradoxes in Big Data (I): Law of Large Populations, Big Data Paradox, and the 2016 US Presidential Election. *The Annals of Applied Statistics* 12 (2): 685–726.
- Mercer, Andrew W. Frauke Kreuter, Scott Keeter, et Elizabeth A. Stuart. 2017. Theory and Practice in Nonprobability Surveys: Parallels Between Causal Inference and Survey Inference. *Public Opinion Quarterly* 81 (S1): 250–71.

- Olson, Kristen, James Wagner, et Raeda Anderson. 2021. Survey Costs: Where are We and What is the Way Forward? *Journal of Survey Statistics and Methodology* 9 (5): 921–42.
- Pasek, Josh. 2018. « Package ‘anesrake’. The Comprehensive R Archive Network. <<https://cran.r-project.org/web/packages/anesrake/anesrake.pdf>>.
- Pasek, Josh, et Jon A. Krosnick. 2010. Measuring intent to participate and participation in the 2010 census and their correlates and trends: Comparisons of RDD telephone and non-probability sample Internet survey data. Statistical Research Division, US Census Bureau, Washington D.C. *Survey Methodology* 2010(15). <<https://www.census.gov/library/working-papers/2010/adrm/ssm2010-15.html>>.
- Kennedy, Courteney, Dana Popky, and Scott Keeter. 2023. How Public Polling Has Changed in the 21st Century. April 19. Pew Research Center.
- Rosenbaum, Paul R. 2005. Sensitivity Analysis in Observational Studies. *Encyclopedia of Statistics in Behavioral Science* 4: 1809–14.
- Rubin, Donald B. 1978. Bayesian Inference for Causal Effects: The Role of Randomization. *The Annals of Statistics* 6 (1): 34–58.
- Smith, Terence M. F. 1991. Post-Stratification. *Journal of the Royal Statistical Society Series D: The Statistician* 40 (3): 315–23.
- Stern, Michael J., Ipek Bilgen, Colleen McClain, et Brian Hunscher. 2017. Effective Sampling from Social Media Sites and Search Engines for Web Surveys: Demographic and Data Quality Differences in Surveys of Google and Facebook Users. *Social Science Computer Review* 35 (6): 713–32.
- Valliant, Richard, et Jill A. Dever. 2011. Estimating Propensity Adjustments for Volunteer Web Surveys. *Sociological Methods & Research* 40 (1): 105–37.
- Wang, Wei, David Rothschild, Sharad Goel, et Andrew Gelman. 2015. Forecasting Elections with Non-Representative Polls. *International Journal of Forecasting* 31 (3): 980–91.
- Weisberg, Herbert F. 2005. *The Total Survey Error Approach*. University of Chicago Press.
- Yeager, David S., Jon A. Krosnick, LinChiat Chang, Harold S. Javitz, Matthew S. Levendusky, Alberto Simpser, et Rui Wang. 2011. Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples. *Public Opinion Quarterly* 75 (4): 709–47.

Pour citer cet article: Poirier, William, Anne-Sophie Charest, Yannick Dufresne, Alexandre Fortier-Chouinard et Nadjim Fréchet. 2025. “Sondages probabilistes et autres créatures mythiques : usages et limites de la calibration des sondages non probabilistes.” *Revue canadienne de science politique* 1–18. <https://doi.org/10.1017/S0008423925100528>