

N° 12-001-X au catalogue
ISSN 1712-5685

Techniques d'enquête

La méthode d'ancrage : estimation des effets de l'intervieweur en l'absence d'attribution imbriquée de l'échantillon

par Michael R. Elliott, Brady T. West, Xinyu Zhang et
Stephanie Coffey

Date de diffusion : le 21 juin 2022



Statistique
Canada

Statistics
Canada

Canada

Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à www.statcan.gc.ca.

Vous pouvez également communiquer avec nous par :

Courriel à infostats@statcan.gc.ca

Téléphone entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros suivants :

- | | |
|---|----------------|
| • Service de renseignements statistiques | 1-800-263-1136 |
| • Service national d'appareils de télécommunications pour les malentendants | 1-800-363-7629 |
| • Télécopieur | 1-514-283-9350 |

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site www.statcan.gc.ca sous « Contactez-nous » > « [Normes de service à la clientèle](#) ».

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population du Canada, les entreprises, les administrations et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques exactes et actuelles.

Publication autorisée par le ministre responsable de Statistique Canada

© Sa Majesté la Reine du chef du Canada, représentée par le ministre de l'Industrie 2022

Tous droits réservés. L'utilisation de la présente publication est assujettie aux modalités de l'[entente de licence ouverte](#) de Statistique Canada.

Une [version HTML](#) est aussi disponible.

This publication is also available in English.

La méthode d'ancrage : estimation des effets de l'intervieweur en l'absence d'attribution imbriquée de l'échantillon

Michael R. Elliott, Brady T. West, Xinyu Zhang et Stephanie Coffey¹

Résumé

Les études méthodologiques des effets des intervieweurs humains sur la qualité des données d'enquête ont longtemps été limitées par une hypothèse critique selon laquelle les intervieweurs d'une enquête donnée sont attribués à des sous-ensembles aléatoires de l'échantillon global plus important (également connu sous le nom d'attribution imbriquée). En l'absence de ce type de conception d'étude, les estimations des effets de l'intervieweur sur les mesures d'intérêt de l'enquête, plutôt que les effets de recrutement ou de mesure spécifiquement introduits par les intervieweurs, peuvent refléter des différences entre les intervieweurs dans les caractéristiques des membres de l'échantillon qui leur sont assignés. Les tentatives précédentes d'approximation de l'attribution imbriquée se sont généralement appuyées sur des modèles de régression pour conditionner les facteurs qui pourraient être liés à l'attribution des intervieweurs. Nous proposons une nouvelle approche pour surmonter ce manque d'attribution imbriquée lors de l'estimation des effets de l'intervieweur. Cette approche, que nous appelons la « méthode d'ancrage », tire avantage des corrélations entre les variables observées qui sont peu susceptibles d'être influencées par les intervieweurs (« ancrés ») et les variables qui peuvent être sujettes aux effets de l'intervieweur, et ce, afin d'éliminer les composantes des corrélations induites par l'intervieweur que l'absence d'attribution imbriquée peut engendrer. Nous tenons compte à la fois des approches fréquentistes et bayésiennes, ces dernières pouvant utiliser des renseignements sur les variances de l'effet de l'intervieweur dans les précédents ensembles de données d'une étude, s'ils sont disponibles. Nous évaluons cette nouvelle méthodologie de manière empirique à l'aide d'une étude par simulation, puis nous illustrons son application au moyen de données d'enquête réelles provenant du *Behavioral Risk Factor Surveillance System* (BRFSS), où les identifiants des intervieweurs sont fournis dans les fichiers de données à grande diffusion. Bien que la méthode que nous proposons partage certaines des limites de l'approche traditionnelle, à savoir le besoin de variables associées au résultat d'intérêt qui sont également exemptes d'erreur de mesure, elle permet d'éviter le besoin d'inférence conditionnelle et présente donc de meilleures qualités inférentielles lorsque l'accent est mis sur les estimations marginales. Elle montre également des signes de réduction supplémentaire de la surestimation des effets plus importants de l'intervieweur par rapport à l'approche traditionnelle.

Mots-clés : Corrélation intra-grappe; corrélation interne; effets de plan de sondage; *Behavioral Risk Factor Surveillance System*.

1. Introduction

Malgré tous les efforts déployés par les organismes d'enquête pour normaliser la formation des intervieweurs qui procèdent à des enquêtes téléphoniques et à des enquêtes sur place (Fowler et Mangione, 1989), de nombreux chercheurs ont montré que les estimations des principaux paramètres de la population ont tendance à varier d'un intervieweur à l'autre (par exemple Groves, 2004; Schnell et Kreuter, 2005; West et Olson, 2010; West et Blom, 2017). Cette variabilité peut être due à des signaux verbaux ou non verbaux envoyés (probablement de façon involontaire) par différents intervieweurs, ou à des caractéristiques démographiques de l'intervieweur qui révèlent ses préférences et ses attentes (West et Blom, 2017). Selon les constatations faites, même les questions factuelles plus simples et les

1. Michael R. Elliott, University of Michigan Institute for Social Research, 426 Thompson St., Ann Arbor, MI 48109, University of Michigan Department of Biostatistics, School of Public Health, 1415 Washington Heights, Ann Arbor, MI 48109. Courriel : mreliot@umich.edu; Brady T. West et Xinyu Zhang, University of Michigan Institute for Social Research, 426 Thompson St., Ann Arbor, MI 48109; Stephanie Coffey, US Census Bureau, 4600 Silver Hill Rd, Suitland-Silver Hill, MD 20746.

questionnaires à remplir soi-même varient d'un intervieweur à l'autre, malgré l'assignation aléatoire des répondants aux intervieweurs (par exemple Kish, 1962; Groves et Magilavy, 1986; O'Muircheartaigh et Campanelli, 1998).

Cette corrélation intra-intervieweur, généralement appelée *effet de l'intervieweur*, réduit l'efficacité des estimations de l'enquête et, compte tenu des coûts fixes de l'enquête, diminue la taille effective des échantillons d'une manière semblable à l'échantillonnage en grappes en raison de la présence d'un effet commun à tous les sujets qui induit une corrélation. Elle peut être conceptualisée en termes statistiques comme un effet aléatoire commun à toutes les observations obtenues par un intervieweur donné, dont la variance est appelée « variance de l'intervieweur ». Il est essentiel de tenir compte de cette variance pour obtenir une inférence statistique correcte. En outre, dans le cadre de la surveillance de la collecte des données, les gestionnaires d'enquête peuvent utiliser des estimations sans biais portant sur les effets de l'intervieweur pour repérer les intervieweurs qui ont des effets extrêmes sur des résultats d'enquête particuliers en temps réel et qui peuvent avoir besoin d'une formation supplémentaire pour freiner leurs comportements déplacés.

Une hypothèse principale dans l'estimation de la variance de l'intervieweur, que ce soit par des modèles à effets aléatoires ou indirectement par l'utilisation d'équations d'estimation généralisées ou de méthodes des séries de Taylor, est l'échantillonnage imbriqué, ou l'assignation aléatoire des cas échantillonnés aux intervieweurs. Ainsi, Schnell et Kreuter (2005) estiment les effets de l'intervieweur lors d'une enquête sur place au cours de laquelle les intervieweurs sont intégrés à des unités primaires d'échantillonnage (UPE) et où les répondants d'une UPE sont attribués de façon aléatoire à un intervieweur, tandis que O'Muircheartaigh et Campanelli (1998) utilisent un modèle recoupé dans un plan d'échantillonnage où les répondants sont attribués de façon aléatoire à des intervieweurs qui ont travaillé dans plusieurs UPE. L'échantillonnage imbriqué permet d'assurer une estimation sans biais de la variance de l'intervieweur en garantissant qu'il n'y a pas de variance « trompeuse » introduite par certains types de répondants plus susceptibles d'être attribués à un intervieweur donné (par exemple des répondants plus âgés associés à des intervieweurs travaillant pendant la journée), tout comme la répartition aléatoire assure une estimation sans biais des effets du traitement dans les essais cliniques. Malheureusement, l'échantillonnage imbriqué est irréalisable sur le plan logistique dans de nombreux plans d'échantillonnage.

Des études récentes sur la variance des intervieweurs ont adopté des approches analytiques ponctuelles pour « ajuster » les effets de covariables sélectionnées susceptibles d'introduire une corrélation trompeuse induite par les intervieweurs en fonction des plans d'échantillonnage non imbriqués (par exemple des covariables décrivant les caractéristiques des zones d'échantillonnage), en affirmant que toute variance restante dans les estimations de l'enquête parmi les intervieweurs est principalement attribuable aux intervieweurs (West et Blom, 2017). Si cette approche peut, en principe, permettre de réduire les corrélations trompeuses entre les intervieweurs et les résultats si de telles covariables sont disponibles, elle présente l'inconvénient de nécessiter une inférence conditionnelle pour la variable d'intérêt substantielle.

Cela est particulièrement problématique si notre objectif est une inférence qui tient correctement compte des effets de l'intervieweur dans l'estimation de la variance sans ajuster de manière inadéquate les covariables qui ne sont d'aucun intérêt. Par exemple, si nous nous intéressons à la moyenne d'une variable d'enquête Y , $E(Y) = \mu$, tout en tenant compte de manière adéquate de la variance supplémentaire introduite par la « corrélation intra-grappe » de plusieurs interviews menées par un seul intervieweur, l'ajustement pour les covariables multiples (X_1, \dots, X_p) donne un estimateur de β_0 sous le modèle $E(Y) = \beta_0 + \sum_{k=1}^p \beta_k x_k$. Il est clair que $\mu \neq \beta_0$ ne peut être utilisé que si $\beta_1 = \dots = \beta_p = 0$ (auquel cas il ne peut y avoir d'ajustement pour les corrélations trompeuses entre les intervieweurs et les résultats), $E(X_1) = \dots = E(X_p) = 0$, ou s'il y a une annulation extrêmement improbable des composantes de régression. (Pour les lecteurs qui connaissent l'inférence causale, cela est quelque peu analogue aux modèles structurels marginaux, qui évitent d'utiliser les facteurs de confusion dans un modèle de régression tout en tenant compte de la confusion, Joffe, Ten Have, Feldman et Kimmel (2004), bien que notre approche soit entièrement basée sur un modèle plutôt qu'assistée par un modèle comme dans les modèles structurels marginaux). Bien que le centrage des covariables puisse garantir la deuxième condition en l'absence d'interactions, cela n'est pas toujours souhaitable ou constaté, et même si cela est faisable, il se peut que les valeurs résiduelles restantes ne présentent pas les caractéristiques de distribution souhaitées. La présente étude a pour objectif de fournir aux spécialistes de la recherche sur les enquêtes un moyen d'estimer la variance de l'intervieweur (soit pour améliorer la qualité des estimations, soit pour informer les opérations des enquêtes) en l'absence d'imbrication sans conditionnement sur les covariables de la manière traditionnelle.

Notre approche, que nous appelons la « méthode d'ancrage », tire avantage des corrélations entre les variables observées qui sont peu susceptibles d'être influencées par les intervieweurs (« ancrés ») et les variables qui peuvent être sujettes aux effets de l'intervieweur (par exemple les questions factuelles complexes ou de nature sensible), et ce, afin d'éliminer statistiquement les composantes des corrélations induites par l'intervieweur qu'une absence d'attribution imbriquée peut engendrer. Les estimations améliorées des effets de l'intervieweur sur les mesures d'enquête augmenteront la capacité des analystes d'enquête à corriger les estimations d'intérêt pour les effets de l'intervieweur et permettront aux gestionnaires d'enquête de gérer de manière adaptative une collecte de données en temps réel et d'intervenir lorsque des intervieweurs particuliers génèrent des résultats d'enquête qui diffèrent considérablement des attentes.

Dans la section 2, nous présentons le contexte de l'important problème de la variance de l'intervieweur, et nous analysons son estimation et de son incidence sur l'inférence. Dans la section 3, nous abordons la méthode d'ancrage et son développement dans un cadre fréquentiste et bayésien, ainsi que l'interprétation heuristique et les questions liées au choix des variables. Dans la section 4, nous évaluons empiriquement les propriétés de cette nouvelle méthode à l'aide d'une étude par simulation, et dans la section 5, nous illustrons la méthode à l'aide de données réelles provenant du *Behavioral Risk*

Factor Surveillance System (BRFSS). Dans la section 6, nous présentons nos conclusions ainsi qu'une discussion sur la mise en œuvre et le suivi de la méthode dans la pratique.

2. Contexte

2.1 Variance de l'intervieweur

La variance entre intervieweurs influe sur les estimations de l'enquête d'une manière semblable aux effets de plan de sondage introduits par l'échantillonnage en grappes. L'augmentation multiplicative de la variance totale d'une moyenne estimée peut être évaluée comme suit : $\text{deff} = 1 + \rho_{\text{int}}(m-1)$, où m est le nombre moyen d'interviews menées par des intervieweurs individuels et ρ_{int} est la corrélation induite par l'intervieweur dans les réponses obtenues à une question d'enquête particulière (Kish, 1965). Des valeurs typiques de 35 répondants par intervieweur et de 0,03 pour ρ_{int} pourraient par conséquent doubler la variance estimée de la moyenne, par rapport à la variance ρ_{int} égale à zéro. Le fait de ne pas tenir compte de la corrélation induite par l'intervieweur introduite par les effets de l'intervieweur conduit à des *effets de spécification incorrecte* (Skinner, Holt et Smith, 1989), ce qui entraîne une inférence non prudente due à la sous-estimation des erreurs-types.

2.2 Estimation de la variance de l'intervieweur

Il est possible que les chercheurs souhaitent estimer la variance de l'intervieweur pour obtenir une inférence statistique correcte (Elliott et West, 2015) et qu'ils souhaitent repérer les intervieweurs ayant des effets inhabituels sur les résultats de la collecte de données aux fins de conception d'enquête réactive, ou comme point central d'une étude méthodologique visant à réduire son incidence en comprenant ses causes (par exemple Brunton-Smith, Sturgis et Williams, 2012; Sakshaug, Tutz et Kreuter, 2013). Les plans imbriqués, qui attribuent les cas échantillonnés aux intervieweurs de façon aléatoire, permettent de tenir compte de la variance des intervieweurs à l'aide de méthodes standards qui tiennent compte de la mise en grappes des données observées, comme les équations d'estimation généralisées (Liang et Zeger, 1986) ou les modèles à effets mixtes (Laird et Ware, 1982; Stiratelli, Laird et Ware, 1984). Si les poids d'échantillonnage sont laissés de côté temporairement, un modèle simple pour une variable d'intérêt normalement distribuée qui tient compte de la variance de l'intervieweur est :

$$Y_{ijk} = \mu + a_i + b_{ij} + \varepsilon_{ijk}, \quad a_i \sim N(0, \sigma_a^2), \quad b_{ij} \sim N(0, \sigma_b^2), \quad \varepsilon_{ijk} \sim N(0, \sigma^2), \quad (2.1)$$

où i indexe une unité primaire d'échantillonnage (UPE), j indexe l'intervieweur dans l'UPE i^e , et k , le répondant, est associé à l'intervieweur j^e dans l'UPE i^e . Si l'on suppose que tous les termes d'erreur sont indépendants, qu'il y a en moyenne J intervieweurs dans chacune des UPE I et qu'il y a en moyenne K interview par intervieweur, la variance de l'estimateur moyen $\hat{\mu} = \bar{y}$ est approximativement gonflée par un facteur de $1 + \rho_a(JK-1) + \rho_b(K-1)$, où $\rho_a = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_b^2 + \sigma^2}$ et $\rho_b = \frac{\sigma_b^2}{\sigma_a^2 + \sigma_b^2 + \sigma^2}$. D'un point de vue pratique,

lorsque la variance $\hat{\mu}$ est la seule quantité d'intérêt, la deuxième étape de corrélation intra-grappe due à un intervieweur peut être négligée, comme dans un plan de « grappe finale » (Kalton, 1983). Lorsque nous traitons l'effet aléatoire de l'UPE comme $\tilde{a}_i = a_i + \sum_{j=1}^J b_{ij}$ avec une variance de $\sigma_{\tilde{a}}^2 = \sigma_a^2 + J\sigma_b^2$, la variance pour l'estimation fondée sur le poids moyen $\hat{\mu}$ est gonflée par un facteur de $1 + \rho_{\tilde{a}}(JK - 1)$, où $\rho_{\tilde{a}} = \frac{\sigma_a^2}{\sigma_{\tilde{a}}^2 + \sigma^2}$.

Si plusieurs intervieweurs sont intégrés dans une seule UPE comme le laisse supposer l'équation (2.1), les variances de l'intervieweur peuvent encore être estimées à des fins méthodologiques à l'aide de modèles linéaires hiérarchiques à plusieurs degrés. Cependant, pour des raisons de rentabilité, de nombreux échantillons probabilistes aréolaires exigent qu'un intervieweur donné limite ses efforts à une seule zone d'échantillonnage (par exemple la National Survey of Family Growth des États-Unis; voir Lepkowski, Mosher, Groves, West, Wagner et Gu, 2013), ce qui déforme complètement les composantes de la variance dues aux intervieweurs et aux zones. De tels plans d'échantillonnage empêchent tout type d'estimation directe de la variance de l'intervieweur, bien que d'un point de vue purement analytique, la prise en compte de la corrélation intra-grappe en utilisant les identifiants des UPE dans l'analyse rendra compte de la variance supplémentaire introduite par l'intervieweur.

Pour d'autres types d'enquêtes, en particulier les enquêtes téléphoniques, cette prise en compte « automatique » des effets de l'intervieweur au stade de l'estimation de la variance que permettent les approches par « grappe finale » n'a pas lieu. Le *Behavioral Risk Factor Surveillance System* (BRFSS; *Centers for Disease Control*, 2013), une enquête téléphonique annuelle massive commanditée par les *Centers for Disease Control*, en est un exemple spectaculaire. Il s'agit de la seule enquête fédérale sur la santé conçue pour fournir des estimations au niveau de l'État, de facteurs de santé clés comme les taux de tabagisme, les mesures de l'obésité et le dépistage du cancer. Elliott et West (2015) n'ont trouvé aucune preuve qu'une proportion substantielle de plus des 1 000 manuscrits publiés à l'aide des données du BRFSS tenait compte des effets de l'intervieweur lors de l'estimation de la variance basée sur ces données, malgré des facteurs d'inflation de la variance de 10 ou plus au niveau de l'État selon des estimations telles que l'autoévaluation de la santé moyenne. Ces auteurs ont trouvé des preuves d'effets substantiels de l'intervieweur pour certains éléments de l'enquête, ainsi qu'une variabilité de la variance de ces effets entre les États, en appliquant à la fois des approches basées sur le modèle et des approches basées sur le plan de sondage pour estimer la variance (bien que cette analyse ait utilisé des estimateurs naïfs contrairement à la régression standard ou aux méthodes d'ancrage dont il est question ici, ce qui pourrait avoir surestimé cette variance).

Il est important de mentionner que les analystes secondaires ne savent toujours pas avec certitude si ces composantes de la variance sont dues à la variabilité de l'échantillonnage, à une véritable erreur de mesure introduite par les intervieweurs ou à une non-réponse différentielle parmi les intervieweurs. En raison de la définition de l'effet de plan de sondage mentionnée ci-dessus, leur incidence sur l'inférence peut quand même être importante même si la corrélation intraclasses (CIC) est faible ou modérée, puisque les intervieweurs mènent généralement de nombreuses interviews. Ainsi, lorsque Groves et Magilavy (1986)

ont trouvé des CIC moyennes comprises entre 0,002 et 0,02 parmi 25 à 55 variables dans chacune des neuf enquêtes téléphoniques portant sur des questions politiques, sanitaires et économiques, l'effet du plan de sondage était compris entre 1,04 et 1,38 pour les études dans lesquelles les intervieweurs réalisent en moyenne 20 interviews chacun, et entre 1,10 et 1,98 si les intervieweurs réalisent en moyenne 50 interviews chacun. Certains résultats peuvent avoir des CIC beaucoup plus élevées : Cernat et Sakshaug (2021) ont trouvé des CIC de l'ordre de 0,30 pour des mesures biométriques, ce qui donnerait des effets de plan de sondage de l'ordre de 15 si 50 interviews étaient menées par intervieweur. Bien que les études sur la variance de l'intervieweur pour les collectes de données en personne aient tendance à être rares parce que les plans d'échantillonnage imbriqués sont plus difficiles à mettre en œuvre dans de tels contextes, Schnell et Kreuter (2005) ont constaté un effet de plan de sondage global médian de 2,0 dans une enquête par sondage à plusieurs étapes sur la crainte suscitée par la criminalité, qui était principalement attribuable aux effets de l'intervieweur plutôt qu'à la mise en grappes spatiales. Pour les analystes, la nécessité de tenir compte des effets de l'intervieweur est donc évidente.

2.3 Prise en compte de la variance de l'intervieweur dans l'inférence en l'absence d'imbrication

Comme nous l'avons mentionné à la section 2.2, lorsque les intervieweurs sont intégrés dans les UPE, les méthodes standards d'estimation de la variance basées sur les « grappes finales » (Kalton, 1983) qui tiennent compte de la dépendance des observations au sein d'une UPE absorberont « automatiquement » l'erreur de mesure due aux intervieweurs dans la corrélation induite par l'UPE. Cependant, lorsque les intervieweurs ne sont pas intégrés dans les UPE, comme cela peut se produire dans certains échantillons probabilistes régionaux où les intervieweurs croisent les segments d'unités d'échantillonnage (par exemple O'Muircheartaigh et Campanelli, 1998; Vassallo, Durrant et Smith, 2017), la corrélation intra-grappe induite par les effets de l'intervieweur doit être prise en compte directement. Dans de telles situations, les modèles à effets aléatoires recoupés (Rasbash et Goldstein, 1994) de la forme

$$E(Y_{hij}) = \theta + a_h + b_i, \quad a_h \sim N(0, \tau_a^2), \quad b_i \sim N(0, \tau_b^2) \quad (2.2)$$

peuvent être employés, lorsque h indexe les UPE, i indexe les intervieweurs, et j indexe les interviews menées par l'intervieweur i^e (par exemple O'Muircheartaigh et Campanelli, 1998; Schnell et Kreuter, 2005; Biemer, 2010; Durrant, Groves, Staetsky et Steele, 2010). Des extensions de ces modèles sont également possibles pour les fonctions de lien non linéaires en utilisant des modèles linéaires mixtes généralisés (par exemple Vassallo et coll., 2017).

Malheureusement, l'imbrication peut échouer, en raison d'une erreur due à la non-réponse différentielle entre les intervieweurs (West et Olson, 2010; West, Kreuter et Jaenichen, 2013), d'une attribution non aléatoire des quarts de travail (par exemple les intervieweurs de jour étant plus susceptibles d'interroger des répondants qui ne travaillent pas) ou d'autres pratiques courantes utilisées pour augmenter les taux de réponse, comme l'attribution d'intervieweurs expérimentés à des répondants plus difficiles

(Brunton-Smith et coll., 2012). En l'absence d'imbrication, les méthodes standards de prise en compte de la variance des intervieweurs peuvent conduire à des corrélations « trompeuses » entre les intervieweurs qui n'ont rien à voir avec l'erreur de mesure induite par l'intervieweur.

La littérature n'est pas totalement dépourvue d'approches permettant d'estimer (et de prendre en compte) la variance de l'intervieweur dans les plans d'échantillonnage non imbriqués. Fellegi (1974), Biemer et Stokes (1985), Kleffe, Prasad et Rao (1991), et Gao et Smith (1998) ont mis au point des méthodes statistiques pour les échantillons aréolaires qui supposent l'imbrication pour un sous-ensemble aléatoire d'UPE, et un seul intervieweur dans chacune des UPE restantes. Des travaux récents ont porté sur les méthodes d'estimation de la variance de l'intervieweur dans les variables d'enquête binaires dans des contextes connexes *d'imbrication partielle* (von Sanden et Steel, 2008). Rohm, Carstensen, Fischer et Gnambs (2021) ont utilisé un modèle de théorie de la réponse d'item à deux paramètres pour séparer les effets de la zone et de l'intervieweur selon cette hypothèse, ce qui permet de supprimer la confusion des effets de l'intervieweur et de la zone dans la mesure où chaque intervieweur recrute dans plusieurs zones et vice versa (bien que l'absence d'assignation aléatoire au sein d'une zone puisse encore produire un certain degré de biais de la composante de variance). Ces méthodes sont utiles pour obtenir des estimations de la variance de l'intervieweur distinctes de l'homogénéité de la zone afin d'évaluer l'incidence indépendante de cette variance. Elles ne sont toutefois pas pertinentes pour notre cadre d'intérêt plus général, où les intervieweurs ne peuvent pas croiser les UPE et ne travaillent pas sur des sous-échantillons aléatoires de l'échantillon complet (c'est-à-dire sans imbrication).

La littérature fait état d'une autre méthode courante utilisée pour faire face au problème des plans d'échantillonnage non imbriqués lors de l'estimation de la variance de l'intervieweur. Cette méthode consiste en l'ajustement des effets des covariables au niveau du répondant et de la zone, ou de l'intervieweur dans les modèles multiniveaux (Hox, 1994; Schaeffer, Dykema et Maynard, 2010; West, Kreuter et Jaenichen, 2013). Ces méthodes sont en grande partie ponctuelles et reposent sur l'hypothèse selon laquelle les covariables incluses rendent compte de manière adéquate de toutes les sources de variabilité qui proviennent des zones (et qui seraient donc attribuées aux intervieweurs si les covariables n'étaient pas prises en compte). Cette approche présente deux inconvénients majeurs. Premièrement, de nombreuses études, en particulier celles qui s'appuient sur des données accessibles au public, peuvent ne pas contenir suffisamment de renseignements sur les covariables au niveau de la zone ou de l'intervieweur pour prendre en compte de manière adéquate le manque de répartition aléatoire dans l'attribution des intervieweurs. Deuxièmement, les estimateurs qui en résultent sont conditionnés par ces covariables, et ces estimateurs conditionnels ne présentent généralement pas d'intérêt, car l'accent est mis sur les estimations marginales des paramètres descriptifs, tels que les moyennes ou les totaux, ou sur les paramètres des modèles qui ne sont généralement pas conditionnés par les covariables (ou qui ne les incluent pas).

3. La méthode d'ancrage

Comme nous l'avons indiqué à la section 2.3, les méthodes existantes ajustent les effets possibles de l'intervieweur introduits aux étapes de recrutement et de mesure de la collecte de données en incluant des covariables au niveau du répondant et de la zone, ou au niveau de l'intervieweur dans les modèles multiniveaux (Hox, 1994). Toutefois, un tel ajustement peut être erroné si une partie de la variance de l'intervieweur est simplement due à un échantillonnage non imbriqué. Comme l'ont indiqué Elliott et West (2015), si des sujets ayant des valeurs semblables à une variable d'intérêt sont assignés à des intervieweurs de manière non aléatoire, par exemple, si un intervieweur travaillant de jour par téléphone a tendance à interroger des répondants plus âgés, dont l'âge peut être corrélé avec les principales variables d'intérêt, ces variables seront corrélées avec des intervieweurs spécifiques. Cependant, nous ne faisons que réordonner l'échantillon aléatoire, sans introduire d'erreur de mesure de la manière décrite dans la section 1, par exemple, West et Blom (2017). Ainsi, les données réelles ne sont pas modifiées et il n'y a pas de véritables effets de l'intervieweur : nous qualifions la corrélation résultante induite par l'intervieweur de « trompeuse » du point de vue de l'inflation de la variance. De cette façon, le fait d'estimer les effets de l'intervieweur sans tenir compte de l'attribution différentielle de l'échantillon peut mener à des inférences prudentes, ce qui entraîne des estimations trompeuses et trop larges de la variance de l'intervieweur, des valeurs p et des intervalles de confiance, ainsi que des décisions opérationnelles incorrectes basées sur les effets prévus pour les intervieweurs individuels.

Pour combler cette importante lacune de la littérature, nous décrivons une méthode « d'ancrage » que les analystes peuvent utiliser pour estimer les composantes uniques de la variance dues aux effets de l'intervieweur sur la sélection et la mesure. La méthode vise à tirer avantage des corrélations entre les variables pour lesquelles l'erreur de mesure de l'intervieweur est préoccupante et les variables qui, selon ce que l'on a raison de croire et ce que l'on sait, sont exemptes d'erreur de mesure. Ainsi, la méthode permet d'éliminer la fraction de la corrélation induite par l'intervieweur qui est due à l'attribution non imbriquée de l'échantillon. Dans le cas le plus simple, si nous avons deux variables, l'une (Y_1) traitée comme étant exempte d'erreur de mesure (l'ancre) et l'autre (Y_2) traitée comme étant susceptible de présenter une erreur de mesure induite par l'intervieweur, et que notre objectif est d'estimer une moyenne de Y_2 , nous ajustons un modèle multiniveau aux données observées pour les deux variables qui inclut un effet aléatoire de l'intervieweur uniquement pour la variable sujette à l'erreur de mesure :

$$y_{ijk} = \mu_k + I(k=2)b_i + \varepsilon_{ijk}. \quad (3.1)$$

Dans (3.1), $i=1, \dots, I$ indexe les intervieweurs, $j=1, \dots, J_i$ indexe les répondants parmi les intervieweurs, $k=1, 2$ indexe la variable (1 = ancre, 2 = variable d'intérêt), $b_i \sim N(0, \sigma_b^2)$ correspond à l'effet de l'intervieweur, et

$$\begin{pmatrix} \varepsilon_{ij1} \\ \varepsilon_{ij2} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \right).$$

Dans ce manuscrit, l'inférence est centrée sur μ_2 , bien que σ_b^2 ou b_i puissent également présenter un intérêt si nous nous concentrons sur la variance des intervieweurs ou sur la détermination des intervieweurs individuels qui contribuent à cette variance.

Afin de fournir une explication heuristique de la raison pour laquelle l'approche d'ancrage proposée fonctionne, supposons que y_{ij1} et y_{ij2} net de b_i sont presque parfaitement corrélés. Étant donné que y_{ij1} est dépourvue d'erreur de mesure, elle peut servir de donnée substitutive de la composante d'erreur non mesurée de y_{ij2} , en absorbant l'erreur artificielle de y_{ij2} induite par le classement des données. L'absence d'imbrication signifie que l'estimation d'un modèle mixte linéaire reposant uniquement sur y_{ij2} donnera une estimation biaisée vers le haut de σ_b^2 si $\sigma_{12} > 0$. Des renseignements seront disponibles pour réduire le biais de $\hat{\sigma}_b^2$. De grands échantillons et des corrélations élevées entre ε_{ij1} et ε_{ij2} donneront des estimations de plus en plus précises de σ_b^2 et donc, de l'incidence réelle de l'erreur de mesure induite par l'intervieweur sur la variance de $\hat{\mu}_2$.

Cette approche peut aisément être généralisée au cas où l'on dispose de variables d'ancrage $K - 1 \geq 2$ exemptes d'erreur de mesure :

$$y_{ijk} = \mu_k + I(k = K) b_{iK0} + \varepsilon_{ijk}, \quad i = 1, \dots, I, j = 1, \dots, J_i, k = 1, \dots, K. \quad (3.2)$$

Dans ce cas, les premières variables $K - 1$ sont supposées être exemptes d'erreur de mesure de la part de l'intervieweur et la variable K^e est la variable d'intérêt, $b_{iK0} \sim N(0, \tau^2)$, et $(\varepsilon_{ij1} \dots \varepsilon_{ijK})^T \sim N_K(0, \Sigma)$, où Σ est une matrice de covariance $K \times K$ non structurée. Sinon, au lieu d'utiliser directement (3.2), il est également possible de ramener (3.2) à une variable bidimensionnelle dans (3.1) en remplaçant Y_{li} par la meilleure prédiction linéaire de Y_{Ki} , en utilisant les variables d'ancrage : $\hat{Y}_{Ki} = E(Y_{Ki} | Y_{1i}, \dots, Y_{K-1,i}) = \hat{\beta}^T \mathbf{X}_i$ où $\hat{\mathbf{X}}_i = (Y_{1i}, \dots, Y_{K-1,i})^T$ et $\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}_K$.

3.1 Remarques sur l'estimation

Il est possible d'utiliser un logiciel standard de modèle linéaire mixte (par exemple SAS PROC MIXED) pour ajuster les modèles dans (3.1) ou (3.2) et obtenir une estimation ponctuelle du maximum de vraisemblance restreint (REML) $\hat{\mu}_2$ ainsi qu'une estimation de la variance connexe. Nous avons fourni un exemple annoté de ce code dans les documents supplémentaires. Les pondérations utilisées pour tenir compte des probabilités inégales de sélection, de l'ajustement pour la non-réponse et de la calibration sur des valeurs de population connues peuvent être incorporées à l'aide de l'estimation du pseudo-maximum de vraisemblance (PML; Pfeiffermann, Skinner, Holmes, Goldstein et Rasbash, 1998; Rabe-Hesketh et Skrondal, 2006) lors de l'ajustement des modèles dans (3.1) ou (3.2). Nous recommandons généralement d'attribuer aux intervieweurs une pondération de 1 lors de l'ajustement de modèles multiniveaux pondérés de ces formulaires, afin d'imiter la notion d'échantillonnage aléatoire simple des intervieweurs à partir d'une population hypothétique d'intervieweurs. Les pondérations des répondants doivent être remises à l'échelle pour que leur somme corresponde au nombre final de répondants pour chaque intervieweur (Carle, 2009), et des extensions de la méthode PML décrites par Veiga, Smith et Brown (2014), et

Heeringa, West et Berglund (2017, chapitre 11) peuvent être utilisées pour incorporer les pondérations remises à l'échelle dans l'estimation de la structure de covariance résiduelle dans (3.1) ou (3.2). Dans les échantillons à plusieurs degrés où les intervieweurs franchissent des zones géographiques, il est également possible d'utiliser des modèles à effets aléatoires recoupés (Rasbash et Goldstein, 1994).

3.2 La méthode d'ancrage bayésienne

En présence de renseignements *a priori* sur les paramètres d'intérêt dans ce modèle (par exemple dans une enquête transversale répétée menée par un intervieweur), les modèles dans (3.1) ou (3.2) peuvent également être ajustés en utilisant une approche bayésienne pour incorporer les renseignements préalables. Dans les enquêtes répétées qui contrôlent soigneusement le rendement de l'intervieweur, il est important d'avoir de bonnes prédictions des effets de l'intervieweur individuel basées sur la composante de variance estimée. En raison des données historiques d'une enquête ayant les mêmes conditions essentielles de conception, il est possible d'estimer les paramètres d'intérêt dans (3.1) en utilisant ces données historiques, et de définir ensuite des distributions *a priori* informatives pour ces paramètres. (Parmi ces types d'enquêtes, citons les collectes de données transversales répétées faisant partie des enquêtes de haute qualité commanditées par les administrations publiques, comme la *National Health Interview Survey* ou, comme l'exemple pris dans cette étude, le *Behavioral Risk Factor Surveillance System*). Plus précisément, nous envisageons l'utilisation d'une distribution *a priori* de l'écart-type de l'effet de l'intervieweur σ_b qui suit une demi-distribution t (Gelman, 2006) ayant des degrés de liberté ν et un écart-type s :

$$p(\sigma_b | \nu, s) = \frac{2\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi s^2}} \left(1 + \frac{\sigma_b^2}{\nu s^2}\right)^{-\frac{\nu+1}{2}}, \quad \tau \geq 0. \quad (3.3)$$

À l'instar de Gelman, nous supposons que $\nu=3$, et nous estimons s en fonction d'estimations *a priori* des effets de l'intervieweur. Nous envisageons des distributions *a priori* standards faibles pour les moyennes à effet fixe : $p(\mu_k) \stackrel{\text{ind}}{\sim} N(0, 10^6)$ et pour la variance résiduelle :

$$p\left(\begin{matrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{matrix}\right) \sim \text{INV - WISHART}(2, I).$$

Cette approche présente des avantages comparativement aux approches de mise à l'essai du rapport de vraisemblance qui reposent sur une théorie asymptotique, en particulier pour les petits échantillons. En utilisant des renseignements *a priori* visant à contraindre la distribution *a posteriori* résultante pour les composantes de la variance de l'intervieweur, nous évitons généralement des tirages extrêmement importants de la composante de la variance, sans pour autant contraindre les moyennes ou les variances résiduelles. Elle contraint également les tirages *a posteriori* des composantes de la variance à être supérieurs à zéro, ce qui permet une inférence basée sur de petites composantes de la variance, alors que les procédures d'ajustement des modèles fréquentistes fixent généralement ces estimations des composantes de la variance exactement à zéro (ce qui équivaut à une hypothèse plutôt déraisonnable selon

laquelle chaque intervieweur produit exactement la même estimation de l'enquête; West et Elliott, 2014). Dans de tels cas, les effets des intervieweurs (même s'ils sont faibles) seraient complètement laissés de côté. L'approche bayésienne permettrait néanmoins d'intégrer les petits effets dans l'inférence. Elle permettrait également d'obtenir des intervalles crédibles pour les composantes de la variance de l'intervieweur en fonction des tirages *a posteriori*.

3.3 Choix des variables d'ancrage

Selon une hypothèse clé sous-jacente de l'approche standard basée sur la régression et de la méthode d'ancrage, les variables sélectionnées sont exemptes d'erreurs induites par l'intervieweur. À l'instar de l'hypothèse de répartition au hasard des données manquantes formulée dans les ouvrages publiés sur les données manquantes, nous ne pensons pas qu'il y aura souvent des cas où nous pourrions être certains de cette hypothèse. Nous croyons plutôt que des approximations pourront être disponibles sur la base de mesures démographiques simples (par exemple l'âge) ou d'autres questions factuelles comportant des choix de réponse simples (par exemple l'emploi actuel), et que l'intervieweur aura peu de chances d'introduire une erreur. L'identification de covariables exemptes d'erreurs avant la collecte des données est une composante substantielle et méthodologique importante de cette approche. La documentation méthodologique *a priori* sur les variables les plus sujettes aux effets de l'intervieweur (West et Blom, 2017) peut être consultée pour cette composante de l'approche.

Comme nous l'avons mentionné ci-dessus, si nous disposons de plusieurs covariables sans erreur mesurées selon les répondants, nous pouvons préserver leur pouvoir prédictif (et donc la corrélation des résidus de l'ancre avec les résidus de la variable d'intérêt) en effectuant un calcul de la prédiction linéaire de la variable d'intérêt à partir d'un modèle linéaire qui inclut les effets fixes de toutes les covariables exemptes d'erreur. Nous envisageons une telle approche dans nos études par simulation et nos applications, et nous la comparons à l'approche « standard », qui consiste simplement en un ajustement de ces covariables dans un modèle multiniveau dans le but d'améliorer l'estimation de la composante de variance de l'intervieweur (Hox, 1994).

Enfin, l'approche d'ancrage s'appuie sur des modèles à effets mixtes susceptibles de produire des estimations correctes avec une quantité suffisante de données. Cependant, ces modèles peuvent être plus difficiles à ajuster, en particulier pour les petits échantillons; par conséquent, nous envisageons également des approches bayésiennes de rechange lors de l'évaluation de l'approche d'ancrage.

4. Étude par simulation

Nous envisageons tout d'abord une étude par simulation empirique de l'approche « d'ancrage » proposée. Nous avons simulé à plusieurs reprises des échantillons de données provenant d'une distribution normale quadrivariée, $(Y_{1ij}^* \ Y_{2ij}^* \ Y_{3ij}^* \ Z_{ij}) \sim N_4(\boldsymbol{\mu}, \Sigma)$, où $j = 1, \dots, J = 30$ indice des répondants hypothétiques intégrés parmi $i = 1, \dots, I = 30$ intervieweurs, $Y_{kij(z)} = Y_{kij(z)}^* + I(k=3)b_i$ pour $b_i \sim N(0, \sigma_b^2)$, et

$Y_{kij(z)}^*$ sont ordonnés par les valeurs de Z_{ij} avant l'attribution des répondants aux 30 intervieweurs. $Y_{1ij(z)}$ et $Y_{2ij(z)}$ sont les valeurs sans erreur de mesure induite par l'intervieweur qui ont été observées, tandis que $Y_{3ij(z)}$ révèle une erreur de mesure induite par l'intervieweur, et Z_{ij} est une covariable (confondante et non observée) qui induit une variabilité externe lorsque le plan de sondage est traité comme étant imbriqué. (Il est possible de considérer Y_1 et Y_2 comme étant des variables démographiques sans erreur de mesure, Y_3 comme étant une mesure continue de la santé globale autodéclarée, qui est potentiellement sujette aux effets de l'intervieweur, et Z comme étant le temps passé à la maison, qui est associé à l'ordonnement de l'intervieweur par quart de travail).

Compte tenu de ce modèle de génération de données, nous notons qu'une corrélation plus élevée de Z avec les autres mesures introduira ce qui semble être la variance de l'intervieweur en raison du classement de $Y_{kij(z)}^*$ par les valeurs de Z_{ij} , au-delà des véritables effets aléatoires de l'intervieweur sur Y_2 (donnés par b_i). Nous souhaitons corriger le manque d'attribution imbriquée avec la méthode d'ancrage proposée, laquelle vise à isoler la variance unique de l'intervieweur σ_b^2 qui ne résulte pas de la simple attribution des cas aux intervieweurs. Pour simplifier, nous supposons que $\mu_{Y_1} = \mu_{Y_2} = \mu_{Y_3} = \mu_z = \mu$, $\sigma_{Y_1}^2 = \sigma_{Y_2}^2 = \sigma_{Y_3}^2 = \sigma_Z^2 = 1$ et $\rho_{Y_1Y_2} = \rho_{Y_1Y_3} = \rho_{Y_1Z} = \rho_{Y_2Y_3} = \rho_{Y_2Z} = \rho_{Y_3Z} = \rho$.

Nous envisageons quatre modèles utilisés pour estimer la moyenne de Y_3 et la variance connexe de l'effet de l'intervieweur :

$$\text{Non ajusté : } Y_{3ij} \sim N(\mu_3 + b_i, \sigma_3^2)$$

$$\text{Ajusté : } Y_{3ij} \sim N(\mu_3 + \beta_1 y_{1ij} + \beta_2 y_{2ij} + b_i, \sigma_3^2)$$

$$\text{Ancrage : } \begin{pmatrix} Y_{1ij} \\ Y_{2ij} \\ Y_{3ij} \end{pmatrix} \sim N_3 \left(\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 + b_i \end{bmatrix}, \Sigma \right)$$

$$\text{Facteur de prédiction linéaire d'ancrage : } \begin{pmatrix} \hat{Y}_{3ij} \\ Y_{3ij} \end{pmatrix} \sim N_2 \left(\begin{bmatrix} \mu_1 \\ \mu_2 + b_i \end{bmatrix}, \Sigma \right)$$

où y_{kij} est la réalisation observée de Y_{kij} , $b_i \sim N(0, \sigma_b^2)$ et, dans le modèle de facteur de prédiction linéaire d'ancrage, $\hat{Y}_{3ij} = \hat{\beta}_0 + \hat{\beta}_1 y_{1ij} + \hat{\beta}_2 y_{2ij}$ où $\hat{\beta}$ est obtenu à partir de la régression linéaire de Y_3 sur Y_1 et Y_2 . Nous estimons la moyenne de Y_3 comme étant l'estimateur REML de μ_3 et, également, la variance connexe de l'effet de l'intervieweur comme étant l'estimateur REML de σ_b^2 .

Nous tenons compte de la possibilité de rejeter l'hypothèse nulle selon laquelle la moyenne des variables observées est nulle (au niveau de 0,05) et le biais empirique dans l'estimation de la variance des effets aléatoires de l'intervieweur, σ_b^2 . Nous avons évalué le biais empirique en calculant la différence entre la moyenne des estimations simulées de la composante de variance et la valeur réelle de la composante de variance propre à un scénario de simulation donné. Notre étude par simulation est conçue selon un plan factoriel complet où $\mu = \{0, 0,5\}$, $\rho = \{0,25; 0,5; 0,75\}$ et $\sigma_b^2 = \{0,1; 0,5; 0,9\}$. Nous avons

généralisé 200 simulations indépendantes pour chacun des 18 classements recoupés de valeurs selon ces paramètres. Le tableau 4.1 présente les résultats de l'étude par simulation.

Tableau 4.1

Résultats de l'étude par simulation empirique. La méthode qui donne le meilleur rendement est en italique (Prenez note que lorsque $\mu = 0$, la puissance idéale est de 0,05)

Valeurs réelles des paramètres du modèle			Puissance : $H_0 : \mu=0$ vs. $H_A : \mu \neq 0$				Biais empirique de $\hat{\sigma}_b^2$			
μ	ρ	σ_b^2	Non ajusté	Ajusté	Ancrage	Facteur de prédiction linéaire d'ancrage	Non ajusté	Ajusté	Ancrage	Facteur de prédiction linéaire d'ancrage
0	0,25	0,1	0,03	<i>0,04</i>	<i>0,04</i>	<i>0,04</i>	0,063	0,029	<i>0,027</i>	<i>0,027</i>
0	0,25	0,5	0,03	0,08	<i>0,04</i>	<i>0,04</i>	0,070	<i>0,022</i>	0,033	0,032
0	0,25	0,9	0,07	<i>0,04</i>	<i>0,06</i>	<i>0,06</i>	0,078	<i>0,037</i>	0,044	0,043
0	0,5	0,1	0,00	<i>0,03</i>	0,02	0,02	0,255	0,061	<i>0,056</i>	<i>0,056</i>
0	0,5	0,5	0,01	<i>0,04</i>	0,03	0,03	0,247	0,058	<i>0,054</i>	<i>0,053</i>
0	0,5	0,9	0,02	<i>0,04</i>	<i>0,04</i>	<i>0,04</i>	0,251	<i>0,049</i>	0,061	0,061
0	0,75	0,1	0,00	<i>0,02</i>	0,01	0,01	0,568	<i>0,074</i>	0,078	0,076
0	0,75	0,5	0,00	0,04	<i>0,05</i>	<i>0,05</i>	0,555	0,098	<i>0,084</i>	<i>0,084</i>
0	0,75	0,9	<i>0,04</i>	<i>0,04</i>	<i>0,06</i>	<i>0,06</i>	0,602	<i>0,099</i>	0,103	0,103
0,5	0,25	0,1	<i>1,00</i>	<i>1,00</i>	<i>1,00</i>	<i>1,00</i>	0,069	<i>0,025</i>	0,032	0,032
0,5	0,25	0,5	<i>0,96</i>	0,68	<i>0,96</i>	<i>0,96</i>	0,072	0,044	<i>0,034</i>	<i>0,034</i>
0,5	0,25	0,9	<i>0,76</i>	0,48	<i>0,75</i>	<i>0,75</i>	0,075	0,040	<i>0,039</i>	<i>0,037</i>
0,5	0,5	0,1	<i>1,00</i>	0,87	<i>1,00</i>	<i>1,00</i>	0,261	0,062	0,062	<i>0,061</i>
0,5	0,5	0,5	0,92	0,44	<i>0,96</i>	<i>0,96</i>	0,269	<i>0,062</i>	0,067	0,067
0,5	0,5	0,9	0,75	0,24	<i>0,80</i>	<i>0,80</i>	0,248	0,068	0,064	<i>0,063</i>
0,5	0,75	0,1	<i>1,00</i>	0,62	<i>1,00</i>	<i>1,00</i>	0,567	0,079	0,078	<i>0,077</i>
0,5	0,75	0,5	0,81	0,27	<i>0,96</i>	<i>0,96</i>	0,507	0,103	<i>0,082</i>	<i>0,082</i>
0,5	0,75	0,9	0,58	0,22	<i>0,70</i>	<i>0,70</i>	0,598	<i>0,100</i>	0,106	0,106

Plusieurs tendances notables se dégagent des résultats de la simulation présentés dans le tableau 4.1. Premièrement, à mesure que les valeurs de ρ augmentent, la méthode d'ancrage produit des réductions plus importantes de la surestimation de la variance de l'intervieweur par rapport au modèle non ajusté. Rappelons que cela était prévu à l'origine, étant donné le classement initial des observations par Z avant l'attribution aux intervieweurs, qui introduit une variance artificielle entre les intervieweurs. De même, comme prévu, l'estimation de la variance de l'intervieweur à l'aide de l'ajustement des covariables est semblable à la méthode d'ancrage lorsque cette variance n'est pas importante, bien qu'il y ait des preuves d'une réduction un peu plus grande du biais lorsque la variance est importante.

En outre, pour les valeurs non nulles de μ , des valeurs plus élevées de ρ entraînent des améliorations plus importantes de la puissance lors de l'utilisation de la méthode d'ancrage par rapport à l'estimateur non ajusté, car une plus grande partie de la variance externe est correctement attribuée. La méthode non ajustée et la méthode d'ancrage donnent toutes deux une puissance supérieure à celle de l'estimateur ajusté, puisque ce dernier est biaisé pour les moyennes non nulles de Y_{1ij} et Y_{2ij} lorsqu'elles sont corrélées avec Y_{3ij} . Des valeurs plus petites de ρ se rapprochent d'une conception imbriquée et, par conséquent, l'approche d'estimation non ajustée ne produit pas de résultats sensiblement différents de ceux de l'approche ajustée ou d'ancrage. Le biais empirique dans l'estimation de σ_b^2 n'est pas lié à la valeur de σ_b^2 , mais est entièrement une fonction de ρ , puisque cela entraîne la corrélation trompeuse induite par

l'intervieweur due à la valeur non observée de Z . Enfin, nous constatons que le remplacement des valeurs réelles de Y_{1ij} et de Y_{2ij} par une mesure agrégée basée sur leur prédiction linéaire de Y_{3ij} donne des résultats pratiquement identiques à leur utilisation directe dans la méthode d'ancrage. Cela est en partie une fonction de leur normalité commune; nous abordons des limites dans la section Discussion ci-dessous.

5. Application au *Behavioral Risk Factor Surveillance System*

Pour illustrer davantage la mise en œuvre de l'approche que nous proposons, nous analysons les données du *Behavioral Risk Factor Surveillance System* (BRFSS; <https://www.cdc.gov/brfss/index.html>) de 2011 et 2012. Le BRFSS est une grande enquête nationale sur la santé aux États-Unis qui est menée par téléphone par les intervieweurs. Il s'agit de l'une des rares enquêtes nationales à fournir aux utilisateurs de données des variables d'identification de l'intervieweur dans les versions de ses ensembles de données à grande diffusion (Elliott et West, 2015). Ces données permettent d'estimer les composantes de la variance de l'intervieweur pour toutes les mesures du BRFSS. Dans la présente étude, nous utilisons uniquement les données des fichiers de données accessibles au public pour ces deux années.

À titre d'illustration, nous nous penchons sur le cas où la variable d'intérêt (Y_2) est l'état de santé perçu (1 = mauvais, ..., 5 = excellent). Nous définissons une variable « d'ancrage » (Y_1) comme étant le facteur de prédiction linéaire de l'état de santé perçu à partir d'un modèle de régression linéaire ajusté par les moindres carrés ordinaires qui inclut l'âge, un indicateur d'obtention d'un diplôme universitaire, un indicateur du fait d'être une femme et un indicateur de race ou d'ethnicité blanche à titre de covariables. Dans le cadre de cette application, voici les trois raisons pour lesquelles nous avons choisi ces covariables au niveau des répondants : 1) nous pensons qu'elles sont susceptibles d'afficher une erreur de mesure différentielle minimale entre les intervieweurs (West et Blom, 2017); 2) elles sont associées à l'attribution des intervieweurs, car les intervieweurs par téléphone ont tendance à faire des quarts de travail à différents moments de la journée, et l'heure de l'entretien est associée à l'âge et au niveau de scolarité (par exemple les répondants plus âgés et les répondants ayant un faible niveau de scolarité sont plus susceptibles d'être interrogés pendant la journée); 3) elles ont également tendance à être corrélées à l'état de santé perçu (Franks, Gold et Fiscella, 2003).

Dans le cadre de l'application, nous comparons également la capacité de la méthode d'ancrage basée sur ce facteur de prédiction linéaire à réduire les estimations des composantes de la variance à celle de la méthode plus « standard » qui est souvent utilisée dans la pratique. Il s'agit simplement d'ajuster ces covariables au niveau du répondant dans un modèle multiniveau, dans le but d'ajuster les effets fixes de ces covariables lors de l'évaluation de la composante de variance de l'intervieweur (Hox, 1994). Voici nos deux remarques sur cette approche, en particulier en ce qui concerne cette application :

1. Le centrage des covariables sur leurs moyennes (qu'elles soient binaires ou continues) est essentiel à cette approche si l'inférence est axée sur la moyenne de Y_2 , car l'absence de centrage entraînera des estimations « conditionnelles » biaisées de la moyenne de cette variable

qui dépendent des valeurs des covariables (plutôt que de la moyenne globale). Le centrage n'est pas pertinent pour la méthode d'ancrage.

2. Dans certains cas, on pourrait s'attendre à ce que les covariables au niveau de l'intervieweur expliquent davantage la variance artificielle de l'intervieweur due à l'attribution non imbriquée que les covariables au niveau du répondant (par exemple les renseignements sociodémographiques au niveau de la région; Hox, 1994; West et Blom, 2017). Cependant, le BRFSS ne fournit pas de covariables au niveau de l'intervieweur.

5.1 Approche fréquentiste

Dans notre analyse, nous avons envisagé à la fois les approches fréquentistes et bayésiennes, et nous avons effectué des analyses distinctes des données du BRFSS de chacun des 50 États et du district de Columbia pour chaque approche. Nous n'avons retenu que les cas présentant des données complètes sur toutes les variables d'analyse d'intérêt afin de garantir une base de cas commune, quel que soit le type d'analyse effectué. Tout d'abord, dans l'approche fréquentiste, nous avons commencé par estimer les moyennes de la santé autodéclarées d'un État donné en supposant que les données sont indépendantes et identiquement distribuées (c'est-à-dire en laissant de côté les effets aléatoires de l'intervieweur) :

$$Y_{ij2} = \mu_2 + \varepsilon_{ij2}, \quad \varepsilon_{ij2} \sim N(0, \sigma_2^2). \quad (5.1)$$

Nous avons ensuite ajusté un modèle à effets mixtes « naïf » comportant des effets aléatoires de l'intervieweur (de la forme vue dans l'équation (2.1), mais sans effets aléatoires de l'UPE, étant donné l'absence d'UPE dans le plan du BRFSS) aux données de santé autodéclarées (en laissant de côté les autres covariables), en supposant une attribution imbriquée de l'échantillon dans chaque État :

$$Y_{ij2} = \mu_2 + b_i + \varepsilon_{ij2}, \quad b_i \sim N(0, \sigma_b^2). \quad (5.2)$$

Nous avons estimé la composante de variance de l'intervieweur en fonction de ce modèle et avons testé la composante de variance pour en vérifier la signification à l'aide d'un test de rapport de vraisemblance basé sur une combinaison (West et Olson, 2010). Nous avons également évalué le rapport entre la variance estimée de la moyenne de l'état de santé autodéclaré en tenant compte naïvement des effets de l'intervieweur et de la variance de la moyenne, et en supposant un échantillonnage aléatoire simple (c'est-à-dire des données indépendantes et identiquement distribuées). Les ouvrages existants font généralement référence à ce rapport, indiqué dans (5.3), comme étant un « effet de l'intervieweur » sur une estimation descriptive particulière :

$$\text{IntEff}_{\text{naive}} = \frac{\text{var}_{\text{naive}}(\hat{\mu}_2)}{\text{var}_{\text{id}}(\hat{\mu}_2)}. \quad (5.3)$$

Ensuite, après avoir ajusté un modèle de régression linéaire à la variable de l'état de santé perçu et après avoir calculé le facteur de prédiction linéaire de l'état de santé perçu sur la base des coefficients

estimés (désigné dans (5.4) par y_{ij1}), nous avons ajusté le modèle dans (3.1) pour mettre en œuvre l'approche d'ancrage :

$$\begin{aligned} y_{ij1} &= \mu_1 + \varepsilon_{ij1} \\ y_{ij2} &= \mu_2 + b_i + \varepsilon_{ij2} \\ b_i &\sim N(0, \sigma_b^2) \\ \begin{pmatrix} \varepsilon_{ij1} \\ \varepsilon_{ij2} \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}\right). \end{aligned} \quad (5.4)$$

Ici $Y_{ij1} = \hat{\beta}_0 + \sum_p \hat{\beta}_p x_{ip}$, où $\hat{\beta}$ est obtenu à partir de la régression linéaire des covariables d'ancrage p (il y en a quatre dans cette application). Nous avons ensuite calculé le même rapport dans (5.3) en fonction de l'approche d'ancrage, où l'on s'attend à ce que l'ancrage réduise le biais dans l'estimation de l'effet de l'intervieweur qui découlerait de l'approche naïve.

Ensuite, nous avons ajusté un modèle représentant l'approche d'ajustement « standard » (Hox, 1994) de la façon suivante :

$$Y_{ij2} = \mu_2 + \sum_p \beta_p x_p + b_i + \varepsilon_{ij2}, \quad b_i \sim N(0, \sigma_b^2). \quad (5.5)$$

Dans (5.5), le x_p représente les covariables centrées au niveau du répondant indexées par p (les mêmes quatre covariables d'ancrage que dans (5.4)), avec les effets fixes correspondants. Nous avons calculé de nouveau le rapport dans (5.3) représentant l'effet estimé de l'enquêteur pour le comparer aux autres approches. Pour rester concentrés sur la réduction potentielle du biais dans l'estimation de l'effet de l'intervieweur, nous avons fait fi des poids d'échantillonnage dans ces analyses.

5.2 Approche bayésienne

Pour l'approche bayésienne, nous avons appliqué les mêmes types d'analyses comparatives afin d'évaluer la méthode d'ancrage, selon que des renseignements *a priori* sur la composante de variance de l'intervieweur provenant du BRFSS 2011 étaient utilisés ou non. Ces renseignements *a priori* proviennent de la mise en œuvre de l'approche d'ancrage avec le même facteur de prédiction linéaire en 2011 afin de déterminer une estimation *a priori* de la composante de variance de l'intervieweur. Dans tous les cas, nous avons supposé des distributions *a priori* non informatives pour les effets fixes (tirés de (3.1), qui définissent les moyennes des deux variables) et les variances et covariances résiduelles dans les modèles.

Nous avons défini une distribution *a priori* informative pour l'écart-type des effets aléatoires de l'intervieweur en utilisant (3.3), où l'écart-type s est donné par l'écart-type estimé des effets aléatoires de l'intervieweur pour le même État en 2011, et nous avons utilisé les distributions *a priori* faibles sur μ et Σ définis dans la section 2.3. Nous avons mis en œuvre l'approche bayésienne en utilisant PROC MCMC dans le logiciel SAS. Des exemples annotés du code utilisé sont disponibles dans les documents supplémentaires.

5.3 Résultats

La figure 5.1 présente quatre nuages de points permettant de comparer les estimations naïves des effets de l'intervieweur sur la moyenne de l'état de santé perçu pour chacun des 50 États et pour le district de Columbia avec les estimations ajustées basées sur la méthode d'ancrage, la méthode d'ajustement « standard » et les deux approches bayésiennes de rechange concernant la mise en œuvre de la méthode d'ancrage. Toutes les estimations des effets de l'intervieweur ont été calculées en utilisant (5.3).

Les tracés varient en fonction des méthodes utilisées pour mettre en œuvre les approches d'estimation. Nous examinons d'abord un tracé des estimations ajustées des effets de l'intervieweur basées sur la méthode d'ancrage par rapport aux estimations naïves des effets de l'intervieweur de (5.3), en utilisant l'approche fréquentiste décrite ci-dessus (figure 5.1a). Le tracé suivant (figure 5.1b) présente les estimations ajustées basées sur l'approche d'ajustement « standard » consistant à inclure les covariables dans un modèle multiniveau. Le troisième tracé (figure 5.1c) représente la première méthode d'ancrage bayésienne avec une distribution *a priori* non informative. Enfin, le quatrième tracé (figure 5.1d) envisage une fois de plus la méthode d'ancrage bayésienne, mais cette fois avec la distribution *a priori* informative susmentionnée, laquelle est basée sur les analyses des données du BRFSS 2011.

En général, nous constatons que la méthode d'ancrage a tendance à réduire les estimations des effets de l'intervieweur, quelle que soit l'approche utilisée. Dans chaque tracé, les points de données situés sous les lignes à 45 degrés indiquent les États pour lesquels une méthode d'ajustement particulière a réduit les estimations des effets de l'intervieweur. En particulier, la méthode d'ajustement « standard » *augmente* plus souvent les estimations des effets de l'intervieweur d'une manière non triviale par rapport à l'approche naïve (figure 5.1b).

Le tableau 5.1 présente les estimations moyennes et les fourchettes des effets de l'intervieweur dans les 50 États et le district de Columbia selon les différentes méthodes. La méthode d'ancrage a eu tendance à réduire les estimations par rapport à la méthode naïve plus souvent que la méthode d'ajustement, 88,2 % et 72,5 % des États ayant vu une réduction des effets estimés de l'intervieweur selon les méthodes d'ancrage fréquentiste et bayésienne informative, respectivement (contre seulement 60,8 % des États selon la méthode d'ajustement). Des éléments de preuve du tableau 5.1 montrent que l'utilisation de renseignements *a priori* est utile lors de l'application de la méthode d'ancrage bayésienne, mais la version fréquentiste de la méthode d'ancrage présente toujours le meilleur rendement global. Dans certains cas, ces réductions de l'effet de l'intervieweur par rapport à l'approche naïve étaient substantielles : cinq des États ont enregistré des réductions de l'effet estimé de l'intervieweur d'au moins 33 %, quel que soit le type de méthode d'ancrage utilisé. Dans certains cas, l'approche d'ancrage a conduit à une légère augmentation des effets estimés de l'intervieweur. Il s'agissait principalement de cas dans lesquels les effets de l'intervieweur étaient très faibles (indiquant ainsi que l'ajustement proposé ne serait pas nécessaire et que toute augmentation des estimations qui en résulterait serait simplement des données bruitées).

Figure 5.1 Nuages de points comparant les estimations naïves et d'ancrage des effets de l'intervieweur pour les 50 États et le district de Columbia, par approche d'estimation (NI = distribution *a priori* non informative; Inf = distribution *a priori* faiblement informative, basée sur des analyses des données du BRFSS de 2011). Les points situés sous les lignes à 45 degrés dans chaque tracé indiquent les États où une méthode d'ajustement particulière a réduit les estimations des effets de l'intervieweur au-dessous de l'estimation naïve.

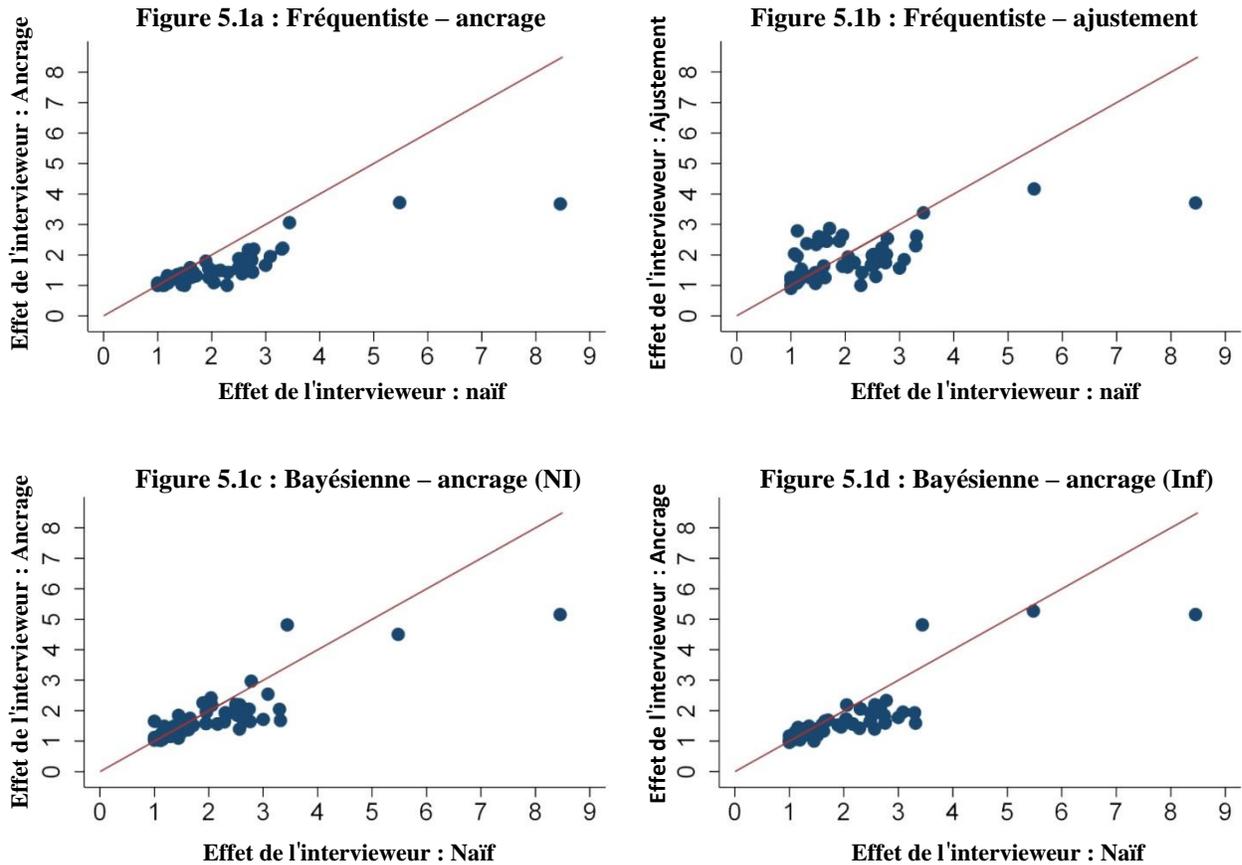


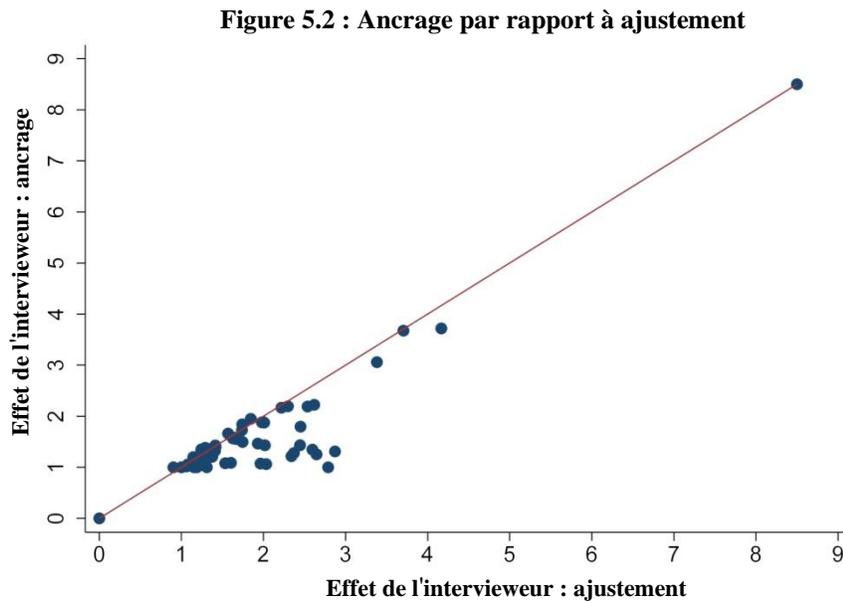
Tableau 5.1

Moyennes et fourchettes des effets de l'intervieweur dans les 50 États et le district de Columbia selon les approches concurrentes

Approche d'estimation	Effets de l'intervieweur : moyenne (fourchette)	Pourcentage d'États ayant une réduction
Fréquentiste – naïve	2,06 (1,00 – 8,45)	-
Fréquentiste – ajustement	1,85 (0,90 – 4,17)	60,8 %
Fréquentiste – ancrage	1,51 (1,00 – 3,72)	88,2 %
Bayésien – ancrage, non informative	1,79 (1,03 – 5,16)	58,8 %
Bayésien – ancrage, informative	1,70 (0,96 – 5,27)	72,5 %

En comparant la méthode d'ancrage à la méthode d'ajustement « standard », nous avons trouvé des preuves cohérentes que la méthode d'ancrage produit des réductions plus importantes des effets estimés de l'intervieweur. La figure 5.2 compare les effets estimés de l'intervieweur pour les 50 États et le district de Columbia lorsque la méthode d'ancrage et la méthode d'ajustement sont utilisées, d'après les résultats fréquentistes seulement. Nous constatons que les estimations des intervieweurs basées sur la méthode d'ajustement ont tendance à être plus importantes que les estimations basées sur l'approche d'ancrage.

Figure 5.2 Nuage de points comparant les estimations ancrées et ajustées des effets de l'intervieweur pour les 50 États et le District de Columbia.



En général, nous n'avons pas trouvé d'avantages importants à utiliser une approche bayésienne pour mettre en œuvre la méthode d'ancrage dans cette application. Nous avons constaté que pour 92,5 % des États, l'intervalle crédible à 95 % pour la composante de variance de l'intervieweur était plus petit en largeur avec l'utilisation de la distribution *a priori* informative qu'avec l'intervalle crédible basé sur la distribution *a priori* non informative, comme il fallait s'y attendre. Cependant, les médianes *a posteriori* des composantes de la variance de l'intervieweur avaient tendance à être similaires selon les deux méthodes d'ancrage bayésiennes (corrélation de Pearson = 0,73).

6. Discussion

Nous avons mis au point et évalué une nouvelle méthode pour estimer les effets de l'intervieweur en l'absence d'attribution imbriquée des unités échantillonnées aux intervieweurs. Grâce à une étude par simulation et à des applications fondées sur des données d'enquête réelles provenant du BRFSS, nous avons démontré la capacité de la méthode d'ancrage proposée à améliorer les estimations des effets de

l'intervieweur dans des situations où l'attribution imbriquée peut ne pas être réalisable, et où la variance de l'intervieweur peut découler des attributions sous-jacentes de l'échantillon. La méthode d'ancrage peut également être facilement appliquée dans un cadre bayésien, en tirant avantage des renseignements *a priori* pour améliorer la qualité des prédictions et des inférences liées aux composantes de variance de l'intervieweur.

Dans les collectes de données d'enquêtes menées par un intervieweur, les effets de l'intervieweur doivent généralement être surveillés dans le cadre d'une collecte de données continue afin d'éviter des problèmes excessifs de variance de l'intervieweur dans les résultats de l'enquête à la fin de la collecte de données. Les gestionnaires d'enquête responsables de ce type de suivi bénéficieront probablement de la méthode d'ancrage et amélioreront ainsi toute décision d'intervention en temps réel prise pour les intervieweurs individuels dans un cadre de conception de l'enquête réactif. Les interventions ou les formations d'appoint en temps réel pour les intervieweurs qui ont des effets extrêmes sur les résultats de production ou les variables d'intérêt scientifique qui, en réalité, ne reflètent que les caractéristiques des domaines dans lesquels ils travaillent et non le rendement réel de l'intervieweur, seront au mieux inefficaces et, au pire, pourraient faire en sorte que les intervieweurs qui offrent habituellement un bon rendement soient critiqués de manière injustifiée et qu'ils abandonnent peut-être une étude donnée.

Lors de l'utilisation de la méthode d'ancrage dans la pratique, nous suggérons qu'elle soit décrite comme une méthode qui permet « d'ajuster les estimations des composantes de la variance de l'intervieweur pour tenir compte de la corrélation trompeuse induite par l'intervieweur dans les mesures d'intérêt de l'enquête qui peuvent survenir en raison de l'attribution non aléatoire des unités échantillonnées aux intervieweurs [traduction] ». Nous soulignons l'importance d'une sélection théorique solide d'une variable (ou de plusieurs variables) d'ancrage qui possède idéalement les propriétés optimales décrites dans le présent article. En l'absence d'une variable d'ancrage ayant ces propriétés optimales, nous soutenons que l'estimation « propre » de la variance de l'intervieweur dans un plan d'échantillonnage non imbriqué peut tout simplement être impossible, et que les analystes : 1) doivent s'ajuster au plus grand nombre possible de covariables au niveau du répondant, de l'intervieweur et de la région lorsqu'ils tentent d'estimer la variance de l'intervieweur, et 2) signalent les estimations de l'incertitude associée aux composantes de la variance estimée, de préférence en utilisant des approches bayésiennes. Ces étapes permettront d'éviter la surestimation des composantes de la variance de l'intervieweur et l'attribution éventuelle d'une qualité des données inférieure aux intervieweurs qui accomplissent déjà des tâches extrêmement difficiles sur le terrain.

La méthode que nous proposons présente plusieurs limites. La plus importante est sans doute le besoin d'avoir des variables d'ancrage qui ne sont pas sujettes à des erreurs de la part de l'intervieweur et qui font toujours l'objet d'une corrélation étroite avec la variable substantielle d'intérêt. Dans notre exemple du BRFS, nous avons considéré l'âge, le sexe, la race et l'éducation comme des ancrages de l'autoévaluation de la santé. Bien que l'âge soit autodéclaré et donc susceptible d'être sujet à un certain degré d'erreur de mesure (par exemple des âges inférieurs à la réalité ou des âges arrondis), nous ne voyons pas de mécanisme évident par lequel cette erreur serait induite par l'intervieweur, bien que cette

possibilité demeure. Un argument similaire peut être avancé pour les trois autres facteurs, même si la possibilité qu'une erreur de mesure induite par l'intervieweur est légèrement plus forte en raison de questions telles que l'affinité entre les intervieweurs et les répondants (West et Blom, 2017). En outre, l'hypothèse de normalité que nous formulons dans l'étude est très restrictive. Pour traiter ce problème dans notre application, nous avons remplacé le modèle d'ancrage multivarié (3.2) par un modèle qui résumait les ancrages multiples en un facteur de prédiction linéaire que nous avons ensuite utilisé dans le modèle d'ancrage bivarié (3.1). Si ce facteur de prédiction linéaire est effectivement une statistique suffisante dans le cas où toutes les variables d'ancrage sont normales, comme le montre l'étude par simulation, il s'agit plutôt d'une solution ponctuelle lorsque certaines ou la totalité des variables d'ancrage ne sont pas normales, comme c'était le cas dans notre application.

Une solution plus conforme aux principes, lorsqu'une ou plusieurs des composantes de Y sont des variables dichotomiques, consisterait à envisager des extensions telles que les modèles probits à effets aléatoires, en remplaçant y_{ijk} dans (3.1) par une latence y_{ijk}^* , où l'observation $y_{ijk} = I(y_{ijk}^* > 0)$ et la variance $\sigma_k^2 = 1$ concernent l'identifiabilité pour toutes les valeurs de k où y_{ijk} est dichotomique. De manière plus ambitieuse, nous pourrions utiliser un modèle de copule à effets aléatoires gaussien (Wu et de Leon, 2014) pour effectuer des distributions arbitraires de Y^* . Les logiciels standards ne pourront pas prendre en charge de tels modèles, bien que les méthodes qui intègrent les effets aléatoires ou qui reposent sur des approches entièrement bayésiennes puissent être envisagées. Ensuite, bien qu'il soit potentiellement important de prendre en compte d'autres sources d'erreur de mesure dans l'inférence, nous nous concentrons ici sur la variance de l'erreur de mesure introduite par les effets de l'intervieweur et son estimation en l'absence d'imbrication. Enfin, nous remarquons que notre approche, comme celle de la concurrence, repose sur des données observées et ne remplace donc pas une véritable imbrication qui garantit l'élimination de toutes les formes d'attribution non aléatoire (observées et non observées).

En plus d'étendre la méthode d'ancrage au cas des coefficients de régression et des variables non normales, les applications futures doivent également tenir compte des contextes dans lesquels les corrélations des variables d'ancrage et les variables d'intérêt de l'enquête pouvant être sujettes aux effets de l'intervieweur sont au mieux modestes. Notre étude par simulation laisse entendre que de bonnes variables d'ancrage ayant des associations fortes avec les variables d'enquête d'intérêt sont importantes pour que cette méthode soit efficace. Les études futures devraient également être axées sur l'identification de variables d'ancrage saines (comme l'âge, le niveau de scolarité, etc.) qui sont peu susceptibles d'être compromises par les intervieweurs et qui pourraient servir d'ancres utiles dans d'autres applications.

Remerciements

Le financement de la présente étude a été assuré par la subvention numéro 1R01AG058599-01 des National Institutes of Health. Les auteurs tiennent à remercier le rédacteur en chef, le rédacteur associé et les deux réviseurs pour leurs conseils qui ont permis d'améliorer le manuscrit.

Documents supplémentaires

Code SAS dans les différentes approches

Le code SAS ci-dessous peut être utilisé pour mettre en œuvre la méthode d'ancrage en utilisant une approche fréquentiste standard. La mise en œuvre de cette approche nécessite que les données fassent partie d'une structure « longue » ayant deux observations par sujet (correspondant aux deux variables), où la variable X2 est une variable indicatrice pour la variable d'ancrage (1 = l'observation sur Y est l'ancre, 0 = l'observation sur Y est la variable d'intérêt), la variable X1 est un indicateur pour la variable d'intérêt (1 = l'observation sur Y est la variable d'intérêt, 0 = l'observation sur Y est l'ancre), INTVID est l'identifiant de l'intervieweur, et OBS est l'identifiant d'un sujet :

```
proc mixed data=yourlongdata;
  class INTVID;
  model y = x2 / solution;
  random x1 / sub=INTVID;
  repeated / sub=obstype=un r rcorr;
run;
```

Le code SAS ci-dessous peut être utilisé pour ajuster le modèle naïf en utilisant une approche bayésienne et une distribution *a priori* faiblement informative. Cette approche exige que les données soient dans le même format « long » :

```
proc mcmc data=yourlongdata seed=41279 nmc=20000 thin=25;
  where x1 = 1; /* ajustable seulement au modèle de la variable d'intérêt */
  parms B0 S2;
  parms Sigma 1;
  prior B: ~ normal(0, var=1e6); /* la distribution a priori des moyennes */
  prior S2 ~ igamma(0.01, scale = 0.01); /* la distribution a priori non
  informative de la variance résiduelle */
  prior Sigma ~ t(0, sd=0.045, df=3, lower=0); /* la distribution a priori
  informative de l'écart-type des effets de l'intervieweur, selon Gelman
  (2006); l'écart-type de la distribution est l'écart-type estimé des effets
  aléatoires de l'intervieweur, qui vise à contraindre la distribution
  a posteriori */
  random Gamma ~ normal(0, sd=Sigma) subject=INTVID;
  Mu = B0 + Gamma; /* le modèle ayant seulement les effets de l'intervieweur
  comme variable d'intérêt */
  model y ~ normal(Mu, var=S2);
run;
```

Enfin, le code SAS ci-dessous peut être utilisé pour mettre en œuvre le modèle naïf selon une approche bayésienne et une distribution *a priori* faiblement informative. La mise en œuvre de cette approche nécessite que les données fassent partie d'une structure large, ayant une ligne par cas et des identifiants d'intervieweurs (INTVID) :

```

proc mcmc data=yourwidedata seed=41279 nmc=20000 thin=25;
  array y[2] genhlthmdd age10; /* var1=variable d'intérêt, var2=ancre */
  array Mu[2]; /* le vecteur selon deux observations par cas */
  array Cov[2,2]; /* la matrice de la covariance résiduelle */
  array S[2,2]; /* est utilisé pour définir la distribution a priori de la
  covariance */
  array H[2] 0 H1; /* H1 = l'effet fixe du changement de la moyenne de l'ancre
  */
  parms B0 Cov; /* l'ordonnée à l'origine (moyenne de la variable d'intérêt)
  et la matrice de la covariance résiduelle */
  parms H1 0; /* le changement de la moyenne de l'ancre */
  parms Sigma 1;
  prior B: H: ~ normal(0, var=1e6); /* la distribution a priori normale des
  effets fixes */
  prior Cov ~ iwish(2,S); /* la distribution a priori de la matrice de
  covariance résiduelle 2x2 */
  prior Sigma ~ t(0, sd=0.045, df=3, lower=0); /* la distribution a priori
  informative de l'écart-type des effets de l'intervieweur, selon Gelman
  (2006); l'écart-type de la distribution est l'écart-type estimé des effets
  aléatoires de l'intervieweur datant de 2011, qui vise à contraindre la
  distribution a posteriori */
  begincnst;
    call identity(S); /* utilise la matrice d'identité pour définir la
    distribution a priori de la matrice de covariance résiduelle
    (noninformative) */
  endcnst;
  random Gamma ~ normal(0,sd=Sigma) subject=INTVID;
  Mu[1] = B0 + Gamma; /* l'effet de l'intervieweur est seulement applicable à
  la variable d'intérêt */
  Mu[2] = B0 + H1; /* la moyenne de l'ancre (note : ce paramétrage est utilisé
  pour garantir des calculs simples de l'écart-type de la distribution
  a posteriori de B0, selon les effets de l'intervieweur */
  model y ~ mvn(Mu, Cov);
run;

```

Bibliographie

- Biemer, P.P. (2010). Total survey error: Design, implementation, and evaluation. *Public Opinion Quarterly*, 74(5), 817-848.
- Biemer, P.P., et Stokes, S.L. (1985). Optimal design of interviewer variance experiments in complex surveys. *Journal of the American Statistical Association*, 80(389), 158-166.
- Brunton-Smith, I., Sturgis, P. et Williams, J. (2012). Is success in obtaining contact and cooperation correlated with the magnitude of the interviewer variance? *Public Opinion Quarterly*, 76, 265-286.
- Carle, A.C. (2009). Fitting multilevel models in complex survey data with design weights: Recommendations. *BMC Medical Research Methodology*, 9, 49-62.

- Centers for Disease Control (2013). [Behavioral Risk Factor Surveillance System: OVERVIEW: BRFSS 2012](http://www.cdc.gov/brfss/annual_data/2012/pdf/Overview_2012.pdf). Accessible à http://www.cdc.gov/brfss/annual_data/2012/pdf/Overview_2012.pdf.
- Cernat, A., et Sakshaug, J.W. (2021). Interviewer effects in biosocial survey measurements. *Field Methods*, 33, 236-252.
- Durrant, G.B., Groves, R.M., Staetsky, L. et Steele, F. (2010). Effects of interviewer attitudes and behaviors on refusal in household surveys. *Public Opinion Quarterly*, 74, 1-36.
- Elliott, M.R., et West, B.T. (2015). "Clustering by interviewer": A source of variance that is unaccounted for in single-stage health surveys. *American Journal of Epidemiology*, 182, 118-126.
- Fellegi, I.P. (1974). An improved method of estimating the correlated response variance. *Journal of the American Statistical Association*, 69,496-501.
- Fowler, F.J., et Mangione, T.W. (1989). *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*. Newbury Park: Sage.
- Franks, P., Gold, M.R. et Fiscella, K. (2003). Sociodemographics, self-rated health, and mortality in the US. *Social Science & Medicine*, 56, 2505-2514.
- Gao, S., et Smith, T.M.F. (1998). A constrained MINQU estimator of correlated response variance from unbalanced data in complex surveys. *Statistica Sinica*, 8, 1175-1188.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1, 515-533.
- Groves, R.M. (2004). Chapter 8: The interviewer as a source of survey measurement error. *Survey Errors and Survey Costs (2nd Edition)*. New York: Wiley-Interscience.
- Groves, R.M., et Magilavy, L.J. (1986). Measuring and explaining interviewer effects in centralized telephone surveys. *Public Opinion Quarterly*, 50, 251-266.
- Heeringa, S.G., West, B.T. et Berglund, P.A. (2017). *Applied Survey Data Analysis, Second Edition*. Boca Raton, FL: Chapman Hall/CRC Press.
- Hox, J.J. (1994). Hierarchical regression models for interviewer and respondent effects. *Sociological Methods and Research*, 22, 300-318.

- Joffe, M.M., Ten Have, T.R., Feldman, H.I. et Kimmel, S.E. (2004). Model selection, confounder control, and marginal structural models: Review and new applications. *The American Statistician*, 58, 272-279.
- Kalton, G. (1983). *Introduction to Survey Sampling*, Sage Publications: Londres, Royaume-Uni.
- Kish, L. (1962). Studies of interviewer variance for attitudinal variables. *Journal of the American Statistical Association*, 57, 92-115.
- Kish, L. (1965). *Survey Sampling*. New York: John Wiley & Sons, Inc.
- Kleffe, J., Prasad, N.G.N. et Rao, J.N.K. (1991). "Optimal" estimation of correlated response variance under additive models. *Journal of the American Statistical Association*, 86, 144-150.
- Laird, N.M., et Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38, 963-974.
- Lepkowski, J.M., Mosher, W.D., Groves, R.M., West, B.T., Wagner, J. et Gu, H. (2013). Responsive design, weighting, and variance estimation in the 2006-2010 National Survey of Family Growth. National Center for Health Statistics. *Vital Health Stat*, 2(158).
- Liang, K.-Y., et Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.
- O'Muircheartaigh, C.A., et Campanelli, P. (1998). The relative impact of interviewer effects and sample design effects on survey precision. *Journal of the Royal Statistical Society, Series A*, 161, 63-77.
- Pfeffermann, D., Skinner, C.J., Holmes, D.J., Goldstein, H. et Rasbash, J. (1998). Weighting for unequal selection probabilities in multilevel models. *Journal of the Royal Statistical Society, Series B*, 60, 23-40.
- Rabe-Hesketh, S., et Skrondal, A. (2006). Multilevel modelling of complex survey data. *Journal of the Royal Statistical Society-A*, 169, 805-827.
- Rasbash, J., et Goldstein, H. (1994). Efficient analysis of mixed hierarchical and cross-classified random structures using a multilevel model. *Journal of Educational and Behavioral Statistics*, 19, 337-350.
- Rohm, T., Carstensen, C.H., Fischer, L. et Gnamb, T. (2021). Disentangling interviewer and area effects in large-scale educational assessments using cross-classified multilevel item response models. *Journal of Survey Statistics and Methodology*, 9, 722-744.

- Sakshaug, J.W., Tutz, V. et Kreuter, F. (2013). Placement, wording, and interviews: identifying correlates of consent to link survey and administrative data. *Survey Research Methods*, 7, 133-144.
- Schaeffer, N.C., Dykema, J. et Maynard, D.W. (2010). Interviewers and Interviewing. Dans *Handbook of Survey Research, Second Edition* (Éds., J.D. Wright et P.V. Marsden), Bingley, Royaume-Uni : Emerald Group Publishing Limited.
- Schnell, R., et Kreuter, F. (2005). Separating interviewer and sampling-point effects. *Journal of Official Statistics*, 21, 389-410.
- Skinner, C.J., Holt, D. et Smith, T.M.F. (1989). *Analysis of Complex Surveys*. New York: John Wiley & Sons, Inc.
- Stiratelli, R., Laird, N. et Ware, J. (1984). Random effects models for serial observations with binary responses. *Biometrics*, 40, 961-971.
- Vassallo, R., Durrant, G. et Smith, P. (2017). Separating interviewer and area effects by using a cross-classified multilevel logistic model: Simulation findings and implications for survey designs. *Journal of the Royal Statistical Society, Series A*, 180, 531-550.
- Veiga, A., Smith, P.W.F. et Brown, J.J. (2014). The use of sample weights in multivariate multilevel models with an application to income data collected by using a rotating panel survey. *Journal of the Royal Statistical Society (Series C)*, 63, 65-84.
- von Sanden, N., et Steel, D. (2008). Optimal estimation of interviewer effects for binary response variables through partial interpenetration. Centre for Statistical and Survey Methodology, University of Wollongong, Document de travail 04-08.
- West, B.T., et Blom, A.G. (2017). Explaining interviewer effects: A research synthesis. *Journal of Survey Statistics and Methodology*, 5, 175-211.
- West, B.T., et Elliott, M.R. (2014). [Approches fréquentiste et bayésienne pour comparer les composantes de l'écart intervieweurs dans deux groupes d'intervieweurs d'enquête](https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2014002/article/14092-fra.pdf). *Techniques d'enquête*, 40, 2, 183-210. Article accessible à l'adresse <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2014002/article/14092-fra.pdf>.
- West, B.T., et Olson, K. (2010). How much of interviewer variance is really nonresponse error variance? *Public Opinion Quarterly*, 74, 1004-1026.

West, B.T., Kreuter, F. et Jaenichen, U. (2013). “Interviewer” effects in face-to-face surveys: A function of sampling, measurement error or nonresponse? *Journal of Official Statistics*, 29, 277-297.

Wu, B., et de Leon, A.R. (2014). Gaussian copula mixed models for clustered mixed outcomes, with application in developmental toxicology. *Journal of Agricultural, Biological, and Environmental Statistics*, 19, 39-56.