

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

Techniques d'enquête

Stratification fondée sur un modèle de populations de paiements dans les enquêtes au sujet de l'intégrité de Medicare

par Don Edwards, Piaomu Liu et Alexandria Delage

Date de diffusion : le 3 janvier 2024



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 le Roi du chef du Canada, représenté par le ministre de l'Industrie, 2024

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.

Stratification fondée sur un modèle de populations de paiements dans les enquêtes au sujet de l'intégrité de Medicare

Don Edwards, Piaomu Liu et Alexandria Delage¹

Résumé

Lorsqu'un fournisseur de soins de santé de Medicare est soupçonné de fraude liée à la facturation, on isole une population de paiements X versés à ce fournisseur sur une certaine période. Un examinateur médical agréé peut, dans un long processus, établir le trop-payé $Y = X -$ (montant justifié par la preuve) pour chaque paiement. En temps normal, il y aura trop de paiements dans une population pour que chacun soit examiné avec soin, aussi prélève-t-on un échantillon probabiliste. Les trop-payés de cet échantillon servent alors à calculer une borne inférieure de l'intervalle de confiance de 90 % pour le trop-payé total de cette population. La borne correspond au montant exigé en recouvrement auprès du fournisseur. Malheureusement, les méthodes classiques de calcul de cette borne ne permettent parfois pas de dégager le niveau de confiance de 90 %, plus particulièrement lorsqu'on utilise un échantillon stratifié.

Dans le présent document, nous présentons et décrivons 166 échantillons épurés tirés des enquêtes au sujet de l'intégrité de Medicare qui comportent 156 populations de paiements correspondantes. Les 7 588 paires échantillonnées (Y, X) indiquent 1) que les vérifications réalisées au sein de Medicare affichent des taux d'erreur élevés : plus de 76 % des paiements en question sont considérés comme étant des erreurs. Elles indiquent aussi 2) que les configurations de ces échantillons vont dans le sens d'un modèle de mélange « tout ou rien » pour (Y, X) qui est déjà défini dans les études spécialisées. Nous analysons des procédures de test de Monte Carlo fondées sur un modèle pour les plans de sondage de Medicare, ainsi que des méthodes de stratification fondées sur les moments anticipés du modèle. Pour la viabilité (atteinte d'un niveau de confiance de 90 %), nous définissons dans le présent article une nouvelle méthode de stratification qui rivalise avec les meilleures parmi de nombreuses méthodes existantes et qui semble moins sensible au choix de paramètres d'exploitation. Pour ce qui est du recouvrement des trop-payés (ce qui équivaut à une mesure de la précision), la nouvelle méthode se compare aussi aux meilleures parmi les nombreuses méthodes expérimentées. Malheureusement, aucun algorithme de stratification mis à l'essai ne s'est révélé viable pour plus de la moitié environ des 104 populations visées par l'essai.

Mots-clés : Fraude de Medicare; modèle de mélange « tout ou rien »; stratification de Dalenius-Hodges; moments anticipés; répartition de Neyman.

1. Introduction

À en croire le rapport des fiduciaires de 2022 des Centers for Medicare and Medicaid Services (CMS) aux États-Unis (<https://www.cms.gov/files/document/2022-medicare-trustees-report.pdf>), les fonds de l'Hospital Insurance Trust du Medicare Trust Fund seront épuisés en 2028. Soucieux de repousser la date de cet épuisement, les CMS et les assemblées législatives se sont concentrés au cours de la dernière décennie à éliminer les fraudes, à améliorer la qualité et à réduire le coût des soins dans l'ensemble (Huffman, 2021; Salmond et Echevarria, 2017). Les CMS procèdent par tests systématiques des taux d'erreur (Comprehensive Error Rate Testing ou CERT; <https://www.cms.gov/Research-Statistics-Data-and-Systems/Monitoring-Programs/Improper-Payment-Measurement-Programs/CERT>); il s'agit d'un programme d'estimation annuelle des taux d'erreur de paiement du Medicare Fee-for-Service (FFS) dans le cadre duquel on échantillonne les

1. Don Edwards, 6, chemin Carpenter, Tybee Island GA 31328. Courriel : edwards@stat.sc.edu; Piaomu Liu, Département des sciences mathématiques, Université Bentley, 175, rue Forest, Waltham, MA 02452; Alexandria Delage, Palmetto GBA, 17 Technology Circle, Columbia, SC 29203.

demandes pour déterminer si les paiements étaient conformes aux règles de couverture, de codage et de paiement de Medicare. En 2022, le taux de paiements irréguliers a été estimé à 7,46 %, ce qui représente 31,46 milliards de dollars de paiements irréguliers. Il est clair que, même lorsque les efforts sont concertés en vue d'éliminer la fraude, le gaspillage et l'abus dans le programme Medicare FFS, des problèmes en matière d'intégrité subsistent en lien avec des demandes de paiements versés de façon inappropriée, et que ces problèmes perdureront (Clemente, McGrady, Repass, Paul III et Coustasse, 2018).

Aux fins de notre exposé, un « fournisseur » se définit comme toute entité facturant des montants à Medicare, qu'il s'agisse des médecins, des fournisseurs de services à domicile, des hôpitaux, des prestataires de services dans des centres de soins palliatifs, des fournisseurs de produits médicaux durables, d'exploitants de services d'ambulance ou d'autres fournisseurs. Les CMS font appel à des entrepreneurs à plusieurs niveaux pour faire enquête sur les fournisseurs soupçonnés d'abuser du système. Les lignes directrices régissant actuellement le processus d'enquête figurent dans le Medicare Program Integrity Manual (MPIM) au chapitre 8 (<https://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/Downloads/pim83c08.pdf>).

S'il y a lieu de soupçonner qu'un fournisseur facture des montants à Medicare FFS de façon inappropriée, un entrepreneur en intégrité du programme unifié (Unified Program Integrity Contractor ou UPIC), qui constituera souvent la filiale d'une société d'assurances, se charge d'effectuer une enquête par sondage. D'abord, l'UPIC obtient des renseignements détaillés sur toutes les réclamations adressées à Medicare réglées à ce fournisseur pendant une certaine période qui sera habituellement d'un à deux ans, ce qui représente une population de paiements. Une unité d'échantillonnage sera choisie et un échantillon probabiliste sera conçu et prélevé. Un examinateur médical agréé examinera ensuite les preuves à l'appui de chaque paiement de l'échantillon pour obtenir le montant Y de trop-payé :

$$Y = \text{trop-payé} = (X \text{ montant payé}) - (\text{montant justifié par la preuve}).$$

Il convient de mentionner que, comme le montant justifié par la preuve n'est pas négatif, le montant de trop-payés aura pour borne supérieure le montant en paiement. Un examen médical prendra habituellement plusieurs mois dans le cas d'un échantillon de taille modérée comme ceux que nous présentons à la section 2. L'UPIC « extrapole » ensuite à partir des montants de trop-payés : il calcule une borne inférieure de confiance de 90 % pour le paiement total versé en trop au fournisseur pendant la période indiquée. Cette borne inférieure est le montant exigé en recouvrement du fournisseur. (Si l'estimation ponctuelle est de la « haute précision » recherchée, elle pourra servir de montant exigé, mais cela est fort rare.)

Dans de telles enquêtes, l'unité d'échantillonnage est d'ordinaire soit la réclamation réglée, soit toutes les réclamations réglées pour les services à un certain bénéficiaire (correspondant plus précisément à un numéro d'identification de Medicare, auparavant associé à un numéro de réclamation relative à l'assurance-maladie (Health Insurance Claim Number ou HICN)). Les plans de sondage prévoient presque toujours un échantillonnage aléatoire simple ou stratifié, les strates étant déterminées par les montants des paiements. La méthode d'extrapolation est habituellement la méthode approximative de la « moyenne par unité » selon le théorème central limite appliqué à une population finie (Hájek, 1964; Li et Ding, 2017).

Plus précisément, considérons un échantillon aléatoire stratifié à L strates, désignons les tailles de strate par N_l et les tailles d'échantillon par n_l , pour $l = 1, 2, \dots, L$. Soit $N = \sum_{l=1}^L N_l$ et $n = \sum_{l=1}^L n_l$. Pour chaque

l , soit $W_l = N_l/N$ et soit \bar{Y}_l et S_l désignant respectivement la moyenne d'échantillon et l'écart-type des trop-payés de la strate l . L'extrapolation selon le théorème central limite est de la forme :

$$N \sum_{l=1}^L W_l \bar{Y}_l - t_{\nu} N \sqrt{\sum_{l=1}^L W_l^2 \left(\frac{S_l^2}{n_l} \right) \left(1 - \frac{n_l}{N_l} \right)} \quad (1.1)$$

où t_{ν} est le 90^e percentile de la distribution t de Student à ν degrés de liberté. Certains entrepreneurs font intervenir les nombres des degrés de liberté approximatifs au moyen de l'approximation de Satterthwaite (1946), mais certains optent pour des méthodes moins prudentes ($n-L$) ou choisissent des degrés infinis de liberté (c'est-à-dire le percentile de la loi normale centrée réduite). Les degrés de liberté de Satterthwaite sont d'au plus ($n-L$) et représentent donc le plus prudent des trois choix. Dans le cas d'un échantillon aléatoire simple, l'équation (1.1), après suppression de l en indice, se ramène à :

$$N\bar{Y} - t_{n-1} N \left(\frac{S}{\sqrt{n}} \right) \sqrt{1 - \frac{n}{N}}. \quad (1.2)$$

Le fournisseur a le droit d'en appeler des décisions de l'examineur quant aux trop-payés ainsi que des méthodes d'échantillonnage employées par l'UPIC. Dans ce cas, un entrepreneur administratif de Medicare indépendant (Medicare Administrative Contractor ou MAC) soumet à une révision le plan de sondage et les décisions relatives au trop-payé au premier niveau d'appel. Si le fournisseur en appelle ensuite de la décision du MAC, un entrepreneur indépendant qualifié (Qualified Independent Contractor) se charge du deuxième niveau d'appel. Si le fournisseur demeure insatisfait, il peut y avoir appel au troisième niveau et l'affaire sera instruite à une audience en bonne et due forme d'un juge en droit administratif (Administrative Law Judge). D'autres appels demeurent possibles. À tout niveau d'appel, le fournisseur peut présenter de nouveaux documents pour justifier ce qu'a été le règlement de ses demandes.

La section 2 du présent exposé vise à examiner les données brutes de 166 échantillons épurés confiés à un même examinateur du MAC de 2013 à 2020. Ces données motivent le choix d'un modèle statistique pour les trop-payés d'échantillons de Medicare à la section 3. À la section 4, nous examinons des méthodes appropriées de test de Monte Carlo pour nos plans de sondage compte tenu du modèle retenu. À la section 5, nous regardons les méthodes existantes de stratification de populations et proposons une nouvelle catégorie de ces méthodes. À la section 6, nous présentons les résultats d'une étude d'efficacité de Monte Carlo où sont comparées ces méthodes de stratification pour 104 des populations de paiements.

2. Faisons honneur aux données : examen de 166 échantillons de Medicare

Nous récapitulons en chiffres et présentons graphiquement les échantillons épurés dans un ordre chronologique approximatif pour la période allant de 2013 à 2020 à l'aide de nuages de points des montants de trop-payés par rapport aux montants de paiements (<https://drive.google.com/drive/folders/1CIzQKzN4-RIIY38WonSAU7ydlCwb76wh>). Nous invitons le lecteur à prendre quelques minutes et à parcourir les

nuages pour se faire sa propre idée des configurations des données. Les enquêtes que nous résumons ici ont été réalisées par quatre entrepreneurs en intégrité du programme. Dans une proportion de 68 % (113), il s'agissait d'échantillons aléatoires simples et, pour les 32 % restants (53), d'échantillons stratifiés selon le montant du paiement. La taille d'échantillon totale variait de 25 à 159, et 90 % des échantillons étaient formés de 30 à 70 paiements au total; 75 % (124) de ces échantillons concernaient les fournisseurs de services à domicile, 16 % (27) concernaient les prestataires de services de centres de soins palliatifs; les 9 % restants (15) concernaient les médecins, les centres de soins spécialisés, les services d'ambulance, etc. Neuf des échantillons avaient le HICN du bénéficiaire comme unité d'échantillonnage et 149, le paiement réglé; 8 échantillons ont été tirés et analysés comme échantillons en cents (Edwards, Gilliland, Ward-Besser et Lasecki, 2015).

Définissons l'« irrégularité » d'un paiement échantillonné comme le rapport entre le trop-payé et le paiement. La figure 2.1 offre une représentation schématique du classement de 7 588 points échantillon issus des 166 échantillons en quatre zones selon la valeur de l'irrégularité : zone 1 (irrégularité $\geq 0,95$), zone 0 ($0 \leq$ irrégularité $\leq 0,05$), zone négative (irrégularité < 0) et zone partielle. Voici quelques observations :

1. Plus de 76 % des paiements en question ont été jugés avoir été partiellement ou complètement versés par erreur. L'opinion toute faite selon laquelle les taux d'erreur à la vérification sont normalement bas ne vaut pas pour les enquêtes de Medicare.
2. Il y avait une seule valeur négative de trop-payé dans les 7 588 paiements passés en revue.
3. Sauf pour les échantillons en « erreur complète » (le trop-payé est égal au paiement dans 24 échantillons), le montant du trop-payé n'est pas en *relation linéaire* avec le montant du paiement dans tous ces échantillons, bien que les deux variables soient parfois fortement en corrélation (l'importante distinction entre « en corrélation » et « en relation linéaire », comme le démontre de manière convaincante le quatuor d'Anscombe (1973) s'estompe malheureusement parfois dans les études spécialisées et celles menées par certains analystes de données). Quelques échantillons comptaient de nombreux trop-payés partiels, auquel cas le rapport entre trop-payé et paiement est essentiellement sans forme dans le triangle constitué par les zones 0 et 1. Pour la majorité de ces échantillons néanmoins, le rapport est décrit au mieux comme étant approximativement *bilinéaire* alors qu'il comporte un trait horizontal dans la zone 0, un trait à 45° dans la zone 1 et peut-être une poignée de trop-payés partiels entre les deux.
4. Cela est moins visible à l'œil nu, mais rien dans ces données ne prouve que plus le paiement augmente, plus les irrégularités augmentent probablement aussi. La figure 2.2, trop remplie, il faut bien l'admettre, trace une courbe des irrégularités par rapport au montant en paiement exprimé en logarithme décimal sans dégager de tendance monotone évidente. La figure 2.3 montre les résultats d'un ajustement par modèle additif généralisé de l'irrégularité par rapport au paiement en logarithme décimal avec une valeur à l'origine propre à chaque échantillon et une fonction lisse commune pour le rapport entre l'irrégularité et le paiement en logarithme décimal.

Cette figure illustre des bandes simultanées de style Scheffé du niveau de confiance de 95 % pour la fonction de régression réelle; ces bandes comportent un trait horizontal. Nous avons obtenu des résultats analogues en utilisant comme variable explicative les montants en paiement au lieu des paiements en logarithme.

Figure 2.1 Classement des 7 588 paires de trop-payé et de paiement dans les 166 échantillons.

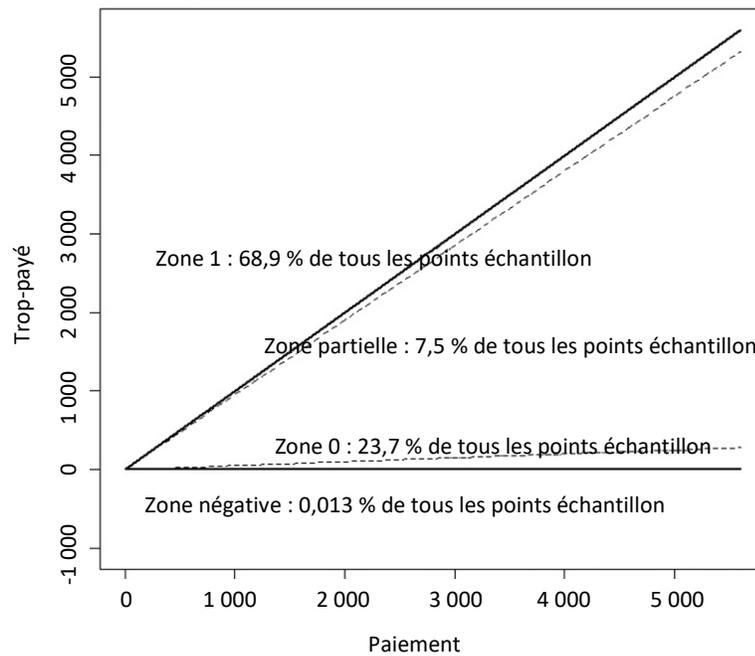


Figure 2.2 Irrégularité par rapport au paiement en logarithme décimal (\log_{10}) pour 7 588 points échantillon.

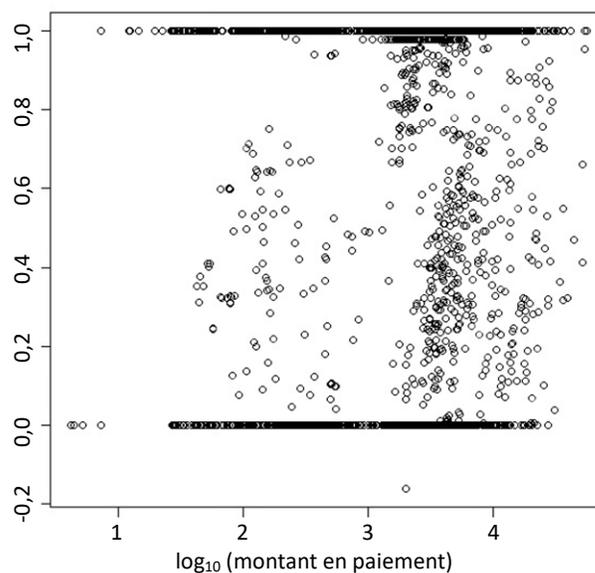
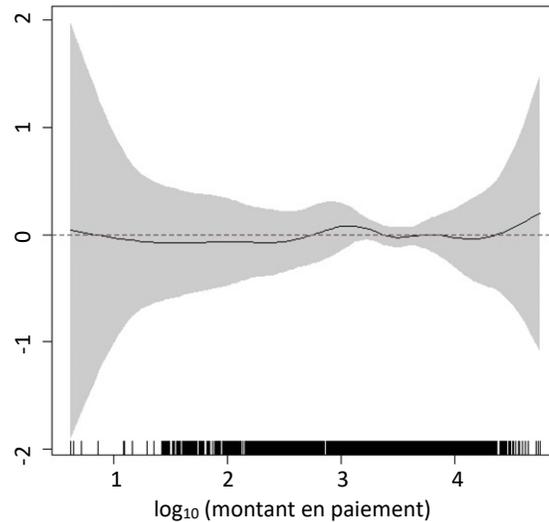


Figure 2.3 Représentation graphique sommaire d'une analyse par modèle additif généralisé de l'irrégularité par rapport au paiement en logarithme décimal (\log_{10}).



Note : Le grisé est une région de style Scheffé de niveau de confiance de 95 % pour la fonction de régression non paramétrique réelle.

3. Un modèle pour les données d'échantillonnage de Medicare

Les configurations des 166 échantillons indiquent un modèle à mélange simple pour les données d'échantillonnage de Medicare. Ce modèle a déjà été examiné par plusieurs auteurs comme King (1996), Edwards, Ward-Besser, Lasecki, Parker, Wieduwilt, Wu et Moorhead (2003) et King et Madansky (2013). Les paiements de population $X_i, i = 1, 2, \dots, N$ sont des constantes connues. Soit $Z_i = 1$ avec la probabilité P_E , $0 \leq P_E \leq 1$, et $Z_i = 0$ autrement, sans égard à la valeur de X_i . Le modèle de mélange « tout ou rien » avec taux d'erreur P_E , ici abrégé en $AN(P_E)$, a comme trop-payé $Y_i = Z_i X_i$. En d'autres termes, la probabilité est P_E que le trop-payé soit égal au paiement et $1 - P_E$ que le trop-payé soit nul. La majorité des 166 échantillons à la section 2 sont conformes à ce modèle, ceux-ci comportant des nombres négligeables de trop-payés partiels ou négatifs.

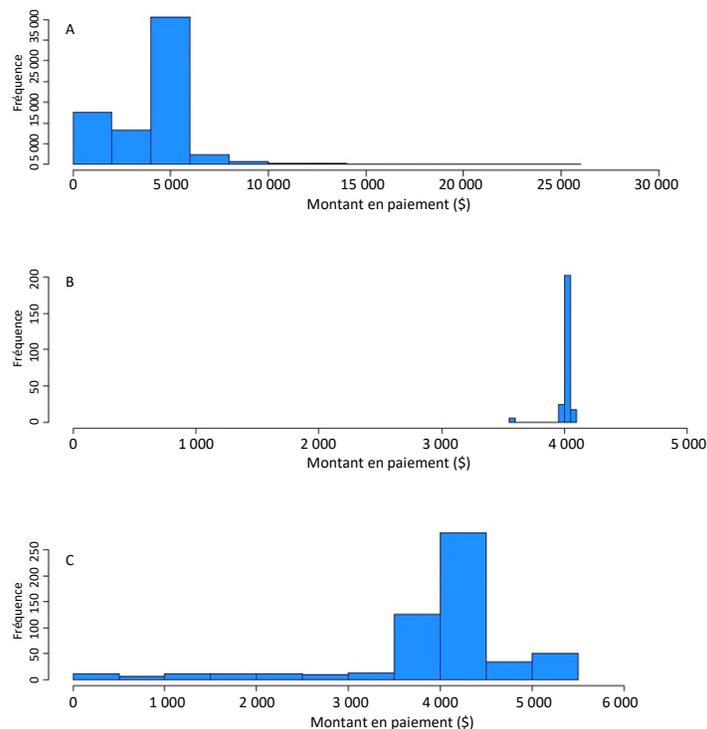
Soit μ_X et σ_X^2 désignant moyenne et la variance connues de la population des montants de paiements. À l'aide du modèle $AN(P_E)$, il est simple de calculer la moyenne et la variance du trop-payé Y pour un X choisi au hasard (King, 1996) :

$$\begin{aligned} E(Y) &= \mu_Y = P_E \mu_X \\ \text{Var}(Y) &= \sigma_Y^2 = \mu_X^2 P_E (1 - P_E) + P_E \sigma_X^2. \end{aligned} \quad (3.1)$$

Il convient de mentionner que la variance des montants de trop-payés peut différer radicalement de la variance des montants des paiements. C'est ce qu'illustre la figure 3.1 qui présente les histogrammes de trois populations de paiements. Nous incluons 0 à l'axe horizontal de ces histogrammes pour mieux nous représenter la forme de la population correspondante de trop-payés pour un taux d'erreur donné P_E , ainsi

que nous allons l'expliquer. Pour la population B et pour des valeurs de P_E proches de $1/2$, la variance des montants de trop-payés σ_Y^2 correspond à plus de 800 fois la variance des montants de paiements σ_X^2 . Dans le cas des populations C et A, le rapport σ_Y^2/σ_X^2 atteint un maximum de 4,7 et 1,5, respectivement. Ainsi, la pratique répandue consistant à employer la variance des montants de paiements comme estimation de la variance des montants de trop-payés dans les formules de détermination selon la taille d'échantillon est inapplicable ici, sauf quand P_E est très proche de 1.

Figure 3.1 Exemple de populations de paiements.



Notes : (A) N = 59 804 paiements relatifs à un HICN pour des services de soins dans des centres de soins palliatifs, (B) N = 249 paiements de réclamations pour des fauteuils roulants électriques; (C) N = 570 paiements de réclamations pour des services de soins dans des centres de soins palliatifs.

4. Test de Monte Carlo des plans de sondage selon le modèle de mélange « tout ou rien »

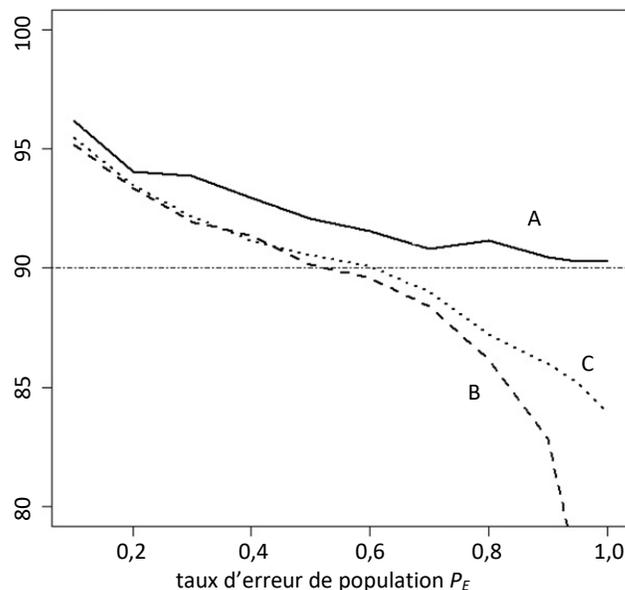
Considérons un plan d'échantillonnage et d'extrapolation proposé pour une population particulière de paiements. Le modèle de tout ou rien $AN(P_E)$ motive le choix d'une procédure simple de test de Monte Carlo pour ce plan à l'aide de la fonction *samptest* du package **R** gratuit (**R** Core Team, 2021); cette fonction peut être obtenue des auteurs. Le test se fait de la manière suivante : pour chaque choix de P_E à une grille couvrant l'intervalle $(0,1]$:

1. on crée une population plausible de trop-payés en choisissant au hasard une proportion P_E pour laquelle le trop-payé est égal au paiement, ce qui fait que la proportion restante $1 - P_E$ a un trop-payé nul;
2. on applique le plan de sondage à la population créée de trop-payés et l'on fait le calcul d'extrapolation et la comparaison avec le trop-payé total de la population.

On reprend indépendamment les étapes 1 et 2 un grand nombre de fois, assez pour que puisse être estimé le niveau de confiance réalisé avec une haute précision. Le plan de sondage visé peut être un échantillonnage aléatoire simple ou stratifié. La fonction résume les tests avec des tracés du niveau de confiance atteint estimé, et le taux moyen de recouvrement des trop-payés pour le plan, par rapport au taux d'erreur P_E (cette dernière quantité correspondant à une précision moyenne, comme nous l'expliquons à la section 6). Pour la plupart des populations de paiements, les résultats de la fonction *samptest* sont renvoyés en quelques secondes même lorsqu'on utilise un ordinateur d'une capacité de calcul modeste.

Ainsi, la figure 4.1 indique les résultats *samptest* de niveau de confiance réalisé pour un échantillon aléatoire simple de 30 paiements lorsqu'une extrapolation standard (1.2) est appliquée aux populations de paiements de la figure 3.1. La courbe nous indique que ce plan de sondage modeste nous donne un niveau de confiance au-dessus ou près du niveau prescrit par le MPIM qui se situe à 90 % pour la population A et tous les taux d'erreur, mais que, dans le cas des populations B et C, il ne donne pas le niveau de confiance recherché de 90 % si le taux d'erreur P_E est de plus de 0,6.

Figure 4.1 Rapport entre le niveau de confiance (%) et le taux d'erreur réalisé pour les populations de la figure 3.1 au moyen d'un échantillon aléatoire simple de 30 paiements et d'une extrapolation type (1.2).



Note : Les estimations présentent une marge de précision de $\pm 1\%$ selon un intervalle de confiance de 95 %.

Ces résultats illustrent le fort effet d'asymétrie de la population de trop-payés sur le niveau de confiance réalisé de la borne (1.2). Cochran (1977, pages 39 à 44) a fait observer qu'une population asymétrique à droite offrait une borne inférieure prudente et une borne supérieure libérale dans un intervalle de confiance bilatéral pour la moyenne de population. À l'aide de ce « principe d'asymétrie » qui, à notre connaissance, est omniprésent, il est possible de prévoir le caractère prudent de la borne inférieure (1.2) pour la population A à tous les taux d'erreur. À un taux d'erreur P_E , la population de trop-payés a un « pic » de zéros d'une taille (approximative) $N(1 - P_E)$, alors que le reste des trop-payés NP_E a la même forme que la population de paiements. Ainsi, la population de trop-payés A est aussi asymétrique à droite pour tout taux d'erreur et, par conséquent, la borne (1.2) est prudente. À un taux d'erreur P_E , la population de trop-payés B a un double pic de zéro de $N(1 - P_E)$ et de paiements de NP_E de près de 4 000 \$. Pour les valeurs de P_E au-dessus de $1/2$, le pic qui se situe à 0 est le plus court des deux, d'où une asymétrie à gauche de la population de trop-payés et une borne (1.2) libérale. Mentionnons enfin que la population de paiements C est asymétrique à gauche. Pour de petites valeurs de P_E , le grand pic qui se situe à 0 dans la population de trop-payés contrebalancera cette asymétrie à gauche et donnera une borne inférieure prudente, mais ce pic qui se situe à 0 disparaît progressivement à mesure qu'augmente P_E , ce qui donne une borne inférieure libérale à des taux d'erreur élevés.

Si des problèmes se posent avec le niveau de confiance réalisé dans le cas de l'extrapolation (1.2), ils se présentent invariablement à des taux d'erreur élevés en raison de l'asymétrie à gauche de la population de trop-payés. C'est la raison pour laquelle la présence de trop-payés partiels vient améliorer le niveau de confiance réalisé : pour tout $P_E > 1/2$ particulier, les trop-payés partiels atténuent toute asymétrie prononcée à gauche. C'est ce qu'a régulièrement confirmé notre expérience de l'utilisation de la fonction *samptest* capable de modéliser avec souplesse le phénomène des trop-payés partiels. Nous supposons que tout plan de sondage qui réalise un niveau de confiance de 90 % au moyen du modèle $AN(P_E)$ pour un taux d'erreur donné P_E le réalisera aussi en présence de trop-payés partiels à cette valeur. En d'autres termes, la présence de trop-payés partiels tend à bonifier le niveau de confiance de la borne (1.2). Cet effet est habituellement modeste, puisque la fréquence de ces trop-payés partiels est elle aussi modeste normalement.

Les populations de paiements asymétriques à droite comme la population A sont fréquentes dans les enquêtes de Medicare. Pour une telle population, un échantillonnage aléatoire simple apportera d'ordinaire un niveau de confiance de plus de 90 % pour tous les taux d'erreur. Ce caractère prudent est l'indice d'une possibilité d'amélioration du plan de sondage par la stratification. On doit cependant être attentif parce que, dans un échantillon stratifié selon le montant des paiements, les effets d'asymétrie sont mixtes. Dans le cas particulier de $L \geq 3$ par exemple, la stratification par montant de paiements crée des sous-populations de paiements d'une forme semblable à celle que présente la figure 3.1B. Cette stratification engendre des strates de trop-payés asymétriques à gauche à des taux d'erreur élevés, ce qui ne permet pas de préserver le niveau de confiance. King et Madansky (2013) le disent bien : [traduction] « Il est fort possible de détruire le niveau de confiance réalisé en stratifiant négligemment. Nous devons souligner le message de mise en garde portant sur le fait qu'une mauvaise application de la stratification (par exemple si l'on n'utilise pas

les limites de strates appropriées) fera plus qu'invalider la précision des bornes; elle pourra invalider aussi le coefficient de confiance normal. » Tout plan de stratification doit être vérifié par un test de Monte Carlo.

Il convient de mentionner que dans certaines situations, une transformation (par exemple la transformation logarithmique) peut normaliser assez bien la distribution pour donner une borne viable de confiance de 90 % pour la moyenne de population de $\log(Y)$, mais que cela ne peut pas faire l'objet d'une rétrotransformation pour obtenir une borne de confiance pour la moyenne de population de Y . En d'autres termes, la moyenne de valeurs logarithmiques n'est pas la valeur logarithmique de la moyenne. Si la transformation rend la distribution symétrique, la borne inférieure découlant de la rétrotransformation sera une borne de confiance pour la médiane de population de Y , et non la moyenne de population. Après multiplication par N , on n'obtient pas une borne pour le total de population, mais quelque chose de bien plus petit. Le même argument vaut pour d'autres transformations non linéaires, qu'il s'agisse de la transformation de la racine carrée ou de la transformation inversée, par exemple.

5. Méthodes de construction de strates à l'aide d'une variable auxiliaire X

Supposons que X est connu pour toutes les unités d'échantillonnage. Dans les enquêtes de Medicare, X est habituellement le montant total payé pour l'unité d'échantillonnage. Nous passerons en revue dans la présente section un certain nombre de méthodes existantes de stratification de population avec X et nous définirons une nouvelle méthode.

Dans certains plans de sondage, quelques montants de paiements parmi les plus élevés sont examinés intégralement dans une strate à tirage complet, ce qui réduit l'ampleur de l'asymétrie à droite dans le reste de la population. Si une telle strate à tirage complet est choisie, il restera à déterminer de quelle façon le reste de la population devrait être stratifié, et ce sera notre propos ici.

Il convient de mentionner au départ qu'il n'y a pas lieu d'employer un algorithme tiré d'une documentation confirmée par les pairs pour déterminer les valeurs de X (« points de démarcation »), $XC_1 < XC_2 < \dots < XC_{L-1}$, définissant les $L \geq 2$ strates. Dans la pratique, tout choix *a priori* permettant de réaliser un niveau de confiance très près ou au-dessus de 90 % pour tous les taux d'erreur est viable, quelle que soit la façon dont il est obtenu. Le temps nécessaire pour trouver ces points de démarcation par tâtonnement en éprouvant chaque choix par le test de Monte Carlo peut cependant être trop élevé. Il est parfois utile de se reporter à un algorithme existant dans la documentation confirmée par les pairs lorsqu'on crée un plan de sondage stratifié.

5.1 Méthodes existantes

Il existe plusieurs méthodes prisées pour établir avec X les points de démarcation ou bornes de stratification. Au moyen de Y comme variable d'intérêt mesurée, Tschuprow (1923) et Neyman (1934) ont

prouvé que la variance de l'estimateur linéaire de moyenne de population $\bar{y}_s = \sum_{l=1}^L W_l \bar{y}_l$ est minimisée pour une taille fixe d'échantillon d'ensemble n , là où la taille d'échantillon n_l dans la strate l est proportionnelle au produit de la taille de la strate N_l par l'écart-type de cette strate $\sigma_{y_l}, l = 1, 2, \dots, L$. Cette « répartition de Neyman » est l'objet même de plusieurs algorithmes de stratification selon diverses hypothèses. La recherche méthodologique dans ce domaine remonte à Dalenius et Gurney (1951) et à Cochran (1977, chapitre 5A). Les études d'intérêt abondent, comme celles de Ekman (1959), Serfling (1968), Singh (1971), Wang et Aggarwal (1984), Hidirolou et Srinath (1993), Hidirolou (1994), Hedlin (2000), Kozak et Verma (2006), Jurina et Gligorova (2017), Hidirolou et Kozak (2018) et Reddy et Khan (2019).

Quand la variable auxiliaire X est en relation linéaire et en haute corrélation avec Y , la méthode bien connue de Dalenius et Hodges (1959) peut permettre une répartition approximative de Neyman. Elle a pour point de départ une distribution de fréquence f pour X . Sous l'hypothèse additionnelle selon laquelle la distribution de X est approximativement uniforme dans les cellules de distribution de fréquences, le choix de bornes de stratification qui égalent le \sqrt{f} cumulatif entre les strates permet d'obtenir une allocation approximative de Neyman pour des tailles d'échantillon de strate égales.

Lavallée et Hidirolou (1988) adoptent un mode itératif de sélection des points de démarcation des strates de manière à minimiser la taille totale de l'échantillon n étant donné un coefficient de variation relatif spécifié pour l'estimateur ponctuel du total de la population; ce problème équivaut à minimiser le coefficient de variation relatif pour un n fixe, ce qui est notre but. Dans cette approche, on suppose que la variable auxiliaire X est « en relation étroite avec » Y . Lavallée et Hidirolou nous avertissent que leur méthode ne sera pas de l'efficacité souhaitée si X et Y ne sont pas « en haute corrélation ». L'algorithme itératif qu'ils utilisent et que nous devons à Sethi (1963) a été amélioré par Kozak (2004).

Pour les plans de sondage où la variable de stratification X n'est pas en relation linéaire avec la variable d'enquête Y , la notion de « moments anticipés » dans un modèle (voir, par exemple Dayal (1985), Sigman et Monsour (1995) et Sweet et Sigman (1995)) joue un rôle fondamental. Généralisant l'algorithme de Lavallée et Hidirolou (1988) en prenant les moments anticipés de Y étant donné X , Rivest (1999, 2002) a proposé des algorithmes et des modèles de stratification qui rendent compte des écarts entre Y et X , plus particulièrement : 1) un modèle loglinéaire (LL) où $\log(Y)$ est en relation linéaire avec $\log(X)$ et 2) un modèle à « remplacement aléatoire ». Ce dernier modélise $Y = X$ avec une forte probabilité, mais autrement Y est égal à une valeur de X choisie au hasard. Les algorithmes de Rivest (1999, 2002) choisissent une taille d'échantillon propre pour réaliser un niveau de précision préétabli ou à maximiser la précision pour obtenir une taille d'échantillon fixe, ce qui permet différentes règles de répartition d'échantillon. Les modèles statistiques et le traitement par moments anticipés dans Rivest (1999, 2002) nous procurent une méthode très générale et très souple permettant de rendre compte d'une grande diversité de scénarios d'enquête.

Baillargeon et Rivest (2009) étendent le modèle LL de Rivest (2002) pour englober une probabilité de survie qui modélise la probabilité que la variable d'enquête Y prenne la valeur 0 (équation (5.1)). On peut

spécifier la probabilité de survie pour qu'elle varie selon les strates. Ce modèle est particulièrement utile dans les enquêtes auprès des entreprises, là où une entreprise n'est plus en activité au moment où l'enquête a lieu, mais où la variable X a été recueillie avec pour résultat une valeur nulle pour Y . Soit p_l la probabilité de survie de la l^e strate; si nous posons $\epsilon \sim N(0, \sigma^2)$, le modèle loglinéaire est le suivant :

$$Y = \begin{cases} \exp(\alpha + \beta \log X + \epsilon), & \text{avec la probabilité } p_l \\ 0, & \text{avec la probabilité } 1 - p_l. \end{cases} \quad (5.1)$$

Dans les enquêtes auprès des entreprises, les valeurs p_l augmentent normalement avec X , ce qui signifie de meilleures chances de survie de l'entreprise, ou avec Y non nul quand la valeur de X est grande. Si une probabilité de survie constante est spécifiée pour toutes les strates, la probabilité que Y soit nul est la même pour toutes les unités d'échantillonnage. Dans ce cas, si $\beta = 0$, $\sigma^2 = 0$, le modèle à l'équation (5.1) équivaut au modèle « tout ou rien » à la section 2 de notre exposé. Cela témoigne de la généralité des modèles proposés par Rivest (1999, 2002) et Baillargeon et Rivest (2009).

Une méthode plus simple d'établissement des points de démarcation d'une stratification pour des populations asymétriques à droite nous vient de Gunning et Horgan (2004) et de Gunning, Horgan et Yancey (2004). Elle est fondée sur l'observation faite par Cochran (1961) que les coefficients de variation sur les différentes strates sont comparables dans une stratification quasi optimale. Si nous supposons que le coefficient de variation est constant à l'échelle des strates, les points de démarcation peuvent se formuler en série géométrique. Une fois le minimum et le maximum de l'ensemble de points de démarcation sont spécifiés, la relation géométrique attribue des points de démarcation pour toutes les strates. Gunning et Horgan déclarent que leur méthode présuppose que Y est « en haute corrélation avec » X . Cette méthode présuppose également que les distributions à l'intérieur des strates sont uniformes en efficacité.

5.2 Répartitions d'échantillon

Pour une taille d'échantillon n ou une cote de précision donnée, les valeurs des points de démarcation subissent l'influence du plan de répartition d'échantillon (Lavallée et Hidiroglou, 1988; Hidiroglou et Srinath, 1993; Horgan, 2006). Les plans de répartition courants comportent notamment (sans s'y limiter) des « répartitions de puissance » comme les répartitions proportionnelles Y et N (Lavallée et Hidiroglou, 1988; Hidiroglou et Kozak, 2018). On peut voir une expression générale des plans de répartition à l'équation (5.2) qui suit, laquelle fait partie du paquetage *stratification* de **R**. Pour la strate l , la taille d'échantillon est $n_l = na_l$. Les combinaisons des paramètres q_1 , q_2 et q_3 produisent diverses répartitions d'échantillon où $0 \leq 2q_1 \leq 1$, $0 \leq 2q_2 \leq 1$ et $0 \leq 2q_3 \leq 1$. N_l , \bar{Y}_l et S_l sont respectivement la taille, la moyenne et l'écart-type de la strate l :

$$a_l = \frac{N_l^{2q_1} \bar{Y}_l^{2q_2} S_l^{2q_3}}{\sum_{l=1}^L N_l^{2q_1} \bar{Y}_l^{2q_2} S_l^{2q_3}}. \quad (5.2)$$

Ainsi, l'emploi de la méthode de Dalenius-Hodges où l'égalité de taille d'échantillon permet de calculer approximativement la répartition de Neyman, ce qui suppose la répartition d'un nombre égal d'observations entre les strates quand nous posons $q_1 = q_2 = q_3 = 0$. Pour une répartition de puissance à 0,7, nous pouvons spécifier $q_1 = q_2 = 0,35$ et $q_3 = 0$. Pour une répartition de puissance proportionnelle Y , la combinaison des paramètres $q_1 = q_3 = 0$ et $q_2 = 0,35$ nous donne une puissance de 0,7.

Dans l'application des méthodes de stratification qui précèdent sous divers plans de répartition d'échantillon, le paquetage *stratification* de **R** (voir Baillargeon et Rivest, 2011) nous offre les fonctions *strata.cumrootf()*, *strata.LH()* et *strata.geo()* pour l'exécution des méthodes de Dalenius-Hodges, de Lavallée-Hidiroglou et de Gunning-Horgan en vue de produire des points de démarcation pour un nombre spécifié de strates. Le paquetage exécute également les autres méthodes étudiées par Rivest (2002) et Baillargeon et Rivest (2009).

5.3 Nouvelle méthode simple utilisant les moments anticipés du modèle « tout ou rien » et des tailles d'échantillon égales

La méthode de stratification bien connue que nous devons à Dalenius et Hodges, sous le modèle $Y \approx X$, choisit de démarcation des strates pour obtenir une quasi-égalité des quantités $N_i \sigma_{Yi}$. Dans ce cas, l'égalité des tailles d'échantillon correspond à la répartition de Neyman. La nouvelle méthode relativement simple qui est décrite dans la présente section imite ce traitement dans le modèle « tout ou rien » avec une valeur particulière P_E . Pour une valeur donnée P_E , l'algorithme $TEE(P_E)$ détermine les points de démarcation en égalisant presque les quantités $N_i \sigma_{Yi}$ déterminées par les moments du modèle « tout ou rien » en (3.1). Ainsi, pour des tailles d'échantillon égales et pour cette valeur de P_E , la méthode $TEE(P_E)$ mène à la répartition de Neyman, ce qui maximise la précision pour un n donné.

Si $L=2$ par exemple, un point de démarcation spécifié XC_1 détermine N_1 et N_2 , ainsi que μ_{1X} , σ_{1X}^2 , μ_{2X} et σ_{2X}^2 . Si nous spécifions P_E , nous pouvons alors utiliser l'équation (3.1) pour calculer σ_{1Y}^2 et σ_{2Y}^2 . Nous dégageons le point de démarcation final par itération sur XC_1 jusqu'à ce que $(N_1 \sigma_{1Y} - N_2 \sigma_{2Y})^2$ soit minimisé. Pour un L général, notre algorithme (disponible auprès du deuxième auteur) procède par recherche non linéaire par la fonction *optimize* de **R** aux fins d'une optimisation unidimensionnelle quand $L=2$, et avec la fonction *constrOptim* pour une optimisation pluridimensionnelle quand $L \geq 3$, ce qui permet de trouver des points de démarcation pour minimiser la somme corrigée des carrés des quantités $N_i \sigma_{Yi}$. Des études exhaustives sont disponibles, mais restent réalisables en calcul seulement pour $L \leq 3$.

La question subsiste : comment choisit-on P_E ? Heureusement et comme nous le verrons à la prochaine section, les caractéristiques opérationnelles de $TEE(P_E)$ semblent plutôt robustes quant au choix qui se fait. Une idée digne d'être étudiée que propose le rédacteur associé est de choisir initialement P_E pour maximiser (3.1). Nous recommandons de sonder les propriétés d'une grille de choix par le test de Monte Carlo avant d'y aller d'un choix définitif.

6. Comparaisons de l'efficacité des méthodes de stratification

Les équations (1.1) et (1.2) peuvent être formulées de façon moins formelle de la manière suivante :

$$(\text{recouvrement des trop-payés}) = (\text{estimation ponctuelle}) - (\text{marge d'erreur}). \quad (6.1)$$

La précision d'un échantillon se définit par sa marge d'erreur exprimée en pourcentage de son estimation ponctuelle; la précision sera tenue pour « élevée » si ce pourcentage est petit. Certains ne définiraient pas la précision comme comprenant le point critique t ; l'essentiel de ce qui est indiqué ci-dessous (à savoir que le taux moyen de recouvrement des trop-payés est en rapport inverse avec la précision moyenne) vaut encore pour les différentes définitions de la précision. Traditionnellement, les analystes s'emploient à créer un plan de sondage d'une haute précision moyenne. Si nous prenons la moyenne (6.1) sur tous les échantillons possibles et considérons que l'estimateur ponctuel est sans biais, nous obtenons ce qui suit :

$$(\text{recouvrement moyen des trop-payés}) = (\text{trop-payé total réel}) - (\text{marge moyenne d'erreur}).$$

Si nous divisons les deux côtés de l'équation qui précède par le trop-payé total réel et multiplions par 100 %, nous obtenons ce qui suit :

$$(\text{taux moyen de recouvrement des trop-payés}) = 100 \% - (\text{précision moyenne}).$$

Ainsi, viser un plan de sondage d'une haute précision moyenne équivaut à chercher un plan ayant un taux de recouvrement moyen de trop-payés élevé. À titre d'exemple, un plan de sondage d'une précision moyenne de 10 % présentera un taux moyen de recouvrement de 90 %. Pour tout plan de sondage, les deux quantités varieront considérablement selon le taux d'erreur P_E .

Nous préférons quantifier l'efficacité d'un plan de sondage par une mesure du recouvrement moyen des trop-payés par opposition à la précision moyenne. Cela facilite la détermination du rapport efficacité-coût pour certains choix de plan de sondage, qu'il s'agisse d'accroître la taille de l'échantillon ou de payer un analyste pour passer une heure de plus à chercher des plans de stratification efficaces. Dans la présente section, nous comparons l'efficacité des principaux algorithmes de stratification présentés à la section 5.

Sur les 166 échantillons de la section 2, 156 populations de paiements, correspondant par numéro, étaient disponibles au total pour être étudiées. Ces populations sont résumées en chiffres et décrites graphiquement à l'adresse <https://drive.google.com/drive/folders/1-7M-4R3KPcCgfPmPWoo24Qc5AOAdOfv7?usp=sharing>. Sur ce nombre, 104 populations de paiements ont été retenues comme candidates à la stratification. Pour ces 104 populations d'essai, des échantillons aléatoires simples (EAS) de taille 30 ont donné un niveau de confiance estimatif à plus de 89 % pour les 7 taux d'erreur vérifiés, $P_E = 0,10, 0,30, 0,50, 0,70, 0,90, 0,95$ et 1,0. Nous avons produit à cette fin 10 000 populations de trop-payés chacune ayant son échantillon et son extrapolation. Nous pouvons supposer sans perte de généralité que, si une strate à tirage complet doit être utilisée, quelques-uns de ces paiements parmi les plus importants auront déjà été écartés pour faire l'objet d'un examen intégral. Les stratégies de stratification ne s'appliquent donc qu'au reste des paiements.

Dans nos essais, des plans de sondage avec une taille totale d'échantillon de 30, 60 ou 90 ont été considérés avec $L = 2, 3$ ou 4 strates pour un total de $104 \times 9 = 936$ cas d'essai. Nous avons mis à l'essai les méthodes de stratification de Dalenius-Hodges (DH), Lavallée-Hidiroglou (LH) et Gunning-Horgan (Geo)

et les modèles loglinéaires (LL) à l'aide du paquetage *stratification* de **R** pour générer des points de démarcation. Trois répartitions d'échantillon ont été testées pour les méthodes LH, Geo et LL : les répartitions de puissance proportionnelles Y et N avec $p = 0,7$ et une répartition puissance avec $q_1 = q_2 = 0,35$ et $q_3 = 0$ sont spécifiées à l'équation (5.2). Nous avons établi les tailles d'échantillon des strates par les fonctions de **R** ayant produit les points de démarcation. Pour les autres méthodes, nous avons tenu pour égales les tailles d'échantillon des strates, sauf dans le cas de $L=4$ et de $n=30$ ou 90 où de légères inégalités apparaissaient.

Pour pleinement définir une méthode de stratification TEE, nous devons choisir une valeur de P_E . Nous pouvons espérer que le choix sera viable et présentera de bonnes propriétés de recouvrement indépendamment de la valeur réelle de P_E , et c'est justement ce que les essais de simulation doivent offrir comme réponse. Dans la présente étude, l'algorithme de points de démarcation TEE a été testé à $P_E = 0,2$ (TEE20), $P_E = 0,5$ (TEE50) et $P_E = 0,8$ (TEE80). Les modèles LL l'ont également été avec des valeurs de probabilité de survie de 0,2, 0,5 et 0,8. L'application du test de Monte Carlo à l'ensemble des cas et des méthodes avec 10 000 itérations à chaque taux d'erreur a demandé environ 27 heures au groupe de calcul de puissance de l'Université Bentley. Les degrés de liberté prudents selon la méthode de Satterthwaite (1946) ont été retenus pour toutes les méthodes.

Pour chacun des 936 cas d'essai, une méthode de stratification était jugée viable si elle réalisait un niveau de confiance estimatif de plus de 89 % pour les sept taux d'erreur vérifiés. Le tableau 6.1 indique, pour chacune des neuf combinaisons de taille d'échantillon et de nombre de strates, le nombre de populations d'essai pour lesquelles chaque méthode a produit un plan de stratification viable. Seuls les résultats de simulation de la répartition de puissance proportionnelle Y sont présentés, car les méthodes de répartition en question ont donné une meilleure viabilité que les deux autres répartitions d'échantillon. Le tableau montre que la stratification est plus susceptible d'être viable pour un L petit et un n grand. La méthode Geo est viable plus souvent que toute autre, mais ce qui peut malheureusement être constaté par ailleurs, c'est qu'elle est habituellement moins efficace qu'un échantillon aléatoire simple de la même taille. Parmi les autres méthodes de stratification, la méthode LL20 et les méthodes TEE20 et TEE50 étaient comparables pour ce qui est de la viabilité et sensiblement meilleures à cet égard que les autres méthodes à l'essai (voir la dernière colonne du tableau 6.1). Fait à noter, ces trois méthodes étaient viables presque deux fois aussi souvent que la méthode de Dalenius-Hodges et presque trois fois aussi souvent que la méthode de Lavallée-Hidiroglou. Il reste que, malheureusement, tout algorithme (sauf le Geo) ne donnait jamais la viabilité pour plus de la moitié environ des populations d'essai.

Dans le but de mesurer l'efficacité (par exemple) de la méthode de Dalenius-Hodges par rapport à l'échantillonnage aléatoire simple en fonction de chacune des tailles d'échantillon, des valeurs L et des populations pour lesquelles la méthode DH était viable, nous avons calculé la différence de recouvrement moyen des trop-payés dans l'ordre DH-EAS pour chaque taux d'erreur. Nous avons calculé des différences analogues pour les autres méthodes de stratification. La figure 6.1 présente des boîtes à moustaches des différences en question. Il ressort de ce graphique que la méthode Geo améliore l'EAS pour moins de la moitié des cas où elle s'est révélée viable. De plus, là où les méthodes DH, LH, LL et TEE étaient viables, elles amélioreraient le recouvrement des trop-payés comparativement à l'EAS d'une valeur médiane de 4 % à 6 %. Dans bon nombre de cas, elles accroissaient considérablement le recouvrement de 10 % à 22 %.

Toutefois, le recouvrement par la méthode DH peut parfois être pire de plus de 9 % par rapport à la méthode EAS. En revanche, la méthode EAS n'a jamais été meilleure de plus de 1,5 % par rapport à n'importe quelle méthode LL, et jamais meilleure de plus de 3 % par rapport à n'importe quelle méthode TEE. Les méthodes et LL et TEE améliorent la méthode EAS jusqu'à 22 %.

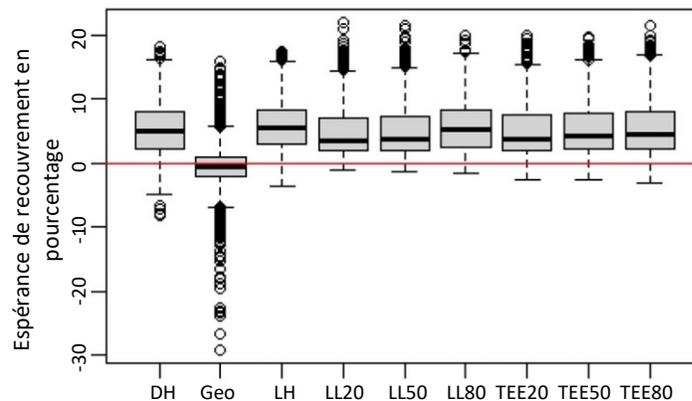
Tableau 6.1

Viabilité des méthodes : nombre de populations d'essai pour lesquelles une méthode de stratification (voir le texte) réalisait un niveau de confiance estimatif d'au moins 89 % pour l'ensemble des taux d'erreur vérifiés par rapport à un nombre de strates L et une taille d'échantillon d'ensemble n dans une répartition de puissance proportionnelle Y à 0,7.

Méthodes	L = 2			L = 3			L = 4			Fréquence de viabilité totale
	n = 30	n = 60	n = 90	n = 30	n = 60	n = 90	n = 30	n = 60	n = 90	
DH	22	30	37	5	8	10	1	3	4	120
LH	19	22	25	4	5	5	0	0	0	80
Geo	90	91	83	52	62	61	5	15	23	482
LL20	31	43	57	19	19	26	9	12	15	231
LL50	28	42	47	14	18	20	5	11	10	195
LL80	21	28	36	7	9	14	2	4	5	126
TEE20	33	41	51	17	22	27	10	13	13	227
TEE50	33	44	45	18	23	24	8	8	12	215
TEE80	30	33	44	11	13	17	5	5	6	164
EAS	104	104	104	104	104	104	104	104	104	936

Notes : DH =Dalenius-Hodges; EAS = Échantillons aléatoires simples; Geo = Gunning-Horgan; LH = Lavallée-Hidiroglou; LL = Loglinéaires; TEE = Tailles d'échantillon égales.

Figure 6.1 Amélioration du recouvrement des trop-payés par rapport à un échantillon aléatoire simple de la même taille pour chaque méthode viable (voir les précisions dans le texte).

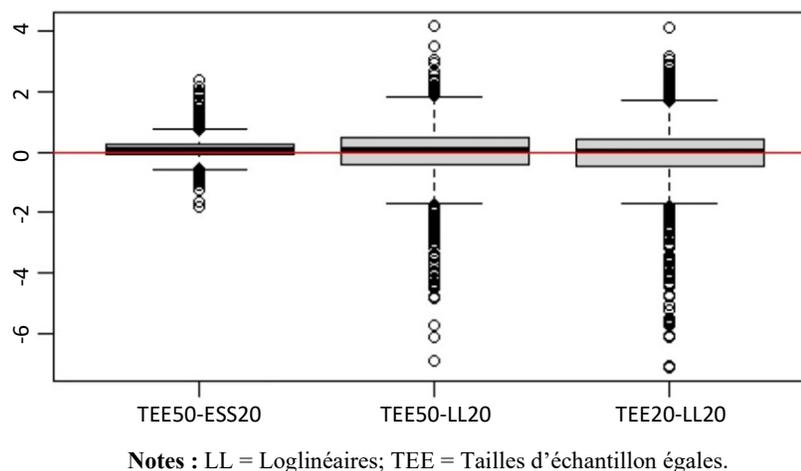


Notes : DH = Dalenius-Hodges; Geo = Gunning-Horgan; LH = Lavallée-Hidiroglou; LL = Loglinéaires; TEE = Tailles d'échantillon égales.

Nous avons ensuite comparé les trois meilleurs algorithmes par calcul des différences dans l'ordre TEE50-TEE20, TEE50-LL20 et TEE20-LL20 chaque fois que les algorithmes appliqués à cette fin étaient viables l'un et l'autre, et ce, pour l'ensemble des taux d'erreur et des cas d'essai. Ces différences sont

présentées à la figure 6.2. Cette figure indique que les méthodes TEE50 et TEE20 étaient d'une même efficacité, d'où l'impression que le choix de P_E dans cette marge n'est pas d'une importance primordiale pour l'efficacité. Dans l'ensemble, les méthodes TEE et LL20 donnaient le même montant de recouvrement des trop-payés prévu. En moyenne, les méthodes TEE produisaient des médianes légèrement supérieures à celles de la méthode LL20. En revanche, la méthode LL20 se caractérisait par une espérance de recouvrement des trop-payés légèrement supérieure en moyenne. Dans certains cas, les méthodes TEE ont surpassé de plus de 4 % la méthode LL20 tandis que dans d'autres cas, elles étaient moins performantes d'un peu plus de 6 % que la méthode LL20.

Figure 6.2 Efficacité relative des méthodes TEE50, TEE20 et LL20 (voir les précisions dans le texte).



7. Analyse et conclusion

La principale réalisation du présent document est une présentation sans précédent des données brutes des enquêtes de Medicare. Il reste que les 166 échantillons indiqués ne peuvent être considérés comme représentatifs de l'ensemble de ces enquêtes. Les échantillons viennent en particulier du premier niveau d'appel; les taux d'erreur tendent à diminuer aux niveaux supérieurs d'appel, les fournisseurs contestant les décisions relatives à la détermination des trop-payés. Ajoutons que les échantillons mentionnés dans le présent document concernent avant tout les fournisseurs de services de soins à domicile ou de services dans des centres de soins palliatifs, ce qui constitue une importante lacune.

Nous ne voyons aucune raison pour laquelle une présentation semblable de données sur des populations de paiements et des échantillons soigneusement épurés ne pourraient pas être faite par l'ensemble des MAC et à des niveaux supérieurs d'appel. Si cela se faisait, disons tous les trois ans, l'anonymat individuel des fournisseurs et des UPIC serait sauvegardé. Étant donné que plus de données seraient communiquées, il serait possible d'encore affiner les modèles du lien entre les trop-payés et les paiements. Il serait également

très utile de savoir si les trop-payés négatifs sont aussi rares et négligeables aux autres niveaux d'appel et pour d'autres fournisseurs de services comme ils pouvaient l'être dans nos échantillons.

Nos données ont indiqué 24 échantillons « en erreur complète », c'est-à-dire des échantillons où chaque trop-payé était égal au paiement. Il peut être difficile en pareil cas d'utiliser la borne (1.1) ou (1.2) fondée sur le théorème central limite de population finie : à des taux d'erreur très élevés, il n'est pas rare que ces bornes inférieures de confiance pour le montant total de trop-payés dépassent le montant total de paiements, ce qui rend leur emploi indéfendable. Les bornes (1.1) et (1.2) peuvent aussi ne pas donner un niveau de confiance d'au moins 90 % pour les taux d'erreur élevés de populations de paiements comme les populations B et C de la figure 3.1. On a mis au point d'autres méthodes d'extrapolation par la distribution probabiliste hypergéométrique (Edwards et coll., 2003; Gilliland et Feng, 2010; Edwards et coll., 2015) pour les situations où les bornes (1.1) et (1.2) laissent à désirer. Ces autres méthodes donnent toujours des bornes inférieures de confiance de 90 % qui sont inférieures au montant total des paiements. Elles sont prudentes mathématiquement – en garantissant un niveau de confiance d'au moins 90 % – tant que les trop-payés négatifs ne se produisent pas plus fréquemment et n'ont pas de conséquences plus graves. Il faudra utiliser plus de données pour vérifier la justesse de cette hypothèse.

Dans le cas des données de Medicare, la fréquence pour laquelle les extrapolations par (1.1) et (1.2) ne sont pas viables est décevante, pour ne pas dire plus. D'autres approches faisant intervenir des bornes de confiance à vraisemblance empirique (Chen, Chen et Rao, 2003; Rao et Wu, 2009) sont prometteuses. Malheureusement, les études de simulation réalisées à ce jour sur les méthodes de vraisemblance empirique ne comportent pas de taux d'erreur de plus de 40 %, ce qui est la règle lorsque les échantillons de Medicare sont utilisés. Une étude systématique des traitements par vraisemblance empirique au moyen des données de Medicare dépasse notre propos, mais elle est au cœur d'un projet en cours des auteurs.

Dans le présent document, nous proposons une nouvelle méthode de stratification comme cas d'espèce de la méthode des moments anticipés dans le modèle « tout ou rien ». L'étude d'efficacité présentée a permis de constater que cette nouvelle méthode TEE pouvait rivaliser avec les meilleures parmi les nombreuses méthodes existantes qui étaient soumises au paquetage *stratification* de **R**. Plus précisément, le modèle loglinéaire avec une probabilité de survie de 0,2 (LL20) dans la répartition de puissance proportionnelle Y ($p = 0,7$) offrait un léger avantage sur le plan de la viabilité par rapport aux meilleures méthodes TEE. Pour trouver ce choix particulier de paramètres d'exploitation dans la vaste gamme disponible dans *stratification*, nous avons néanmoins dû obtenir les conseils d'un expert et prévoir des heures de test. Par contre, les méthodes TEE quasi équivalentes ($P_E = 0,2$ ou $0,5$) comptaient parmi les premières méthodes que nous avons vérifiées. Toutefois, aucune méthode ne dominait l'ensemble des autres méthodes dans toutes les situations. Et pour chaque choix de n et L , il y avait de nombreuses populations d'essai pour lesquelles aucun algorithme de stratification n'était viable.

Comme c'est souvent le cas, ces résultats mènent à de nouvelles questions. Y a-t-il des aspects de la population de paiements qui puissent nous indiquer le moment où la stratification entraînera un gain d'efficacité ? Y a-t-il des aspects de la population de paiements qui puissent nous éclairer sur la méthode

qui représenterait le meilleur choix ? Les auteurs étudieront ces questions et d'autres et ils demanderont la collaboration des autres chercheurs.

Remerciements

Les auteurs remercient les CMS de nous avoir autorisés à utiliser et à présenter les données de ces échantillons et de ces populations. Ils remercient le rédacteur associé et les examinateurs pour leur dur travail et leurs suggestions utiles. Piaomu Liu remercie l'Université Bentley d'avoir fourni des capacités de calcul à haute performance pour l'exécution des études d'efficacité dans ce projet de recherche.

Bibliographie

- Anscombe, F.J. (1973). Graphs in statistical analysis. *The American Statistician*, 27(1), 17-21.
- Baillargeon, S., et Rivest, L.-P. (2009). A general algorithm for univariate stratification. *Revue Internationale de Statistique*, 77(3), 331-344.
- Baillargeon, S., et Rivest, L.-P. (2011). [Élaboration de plans stratifiés en R à l'aide du programme stratification](https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2011001/article/11447-fra.pdf). *Techniques d'enquête*, 37, 1, 59-72. Article accessible à l'adresse <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2011001/article/11447-fra.pdf>.
- Chen, J., Chen, S.-Y. et Rao, J. (2003). Empirical likelihood confidence intervals for the mean of a population containing many zero values. *Canadian Journal of Statistics*, 31(1), 53-68.
- Clemente, S., McGrady, R., Repass, R., Paul III, D.P. et Coustasse, A. (2018). Medicare and the affordable care act: Fraud control efforts and results. *International Journal of Healthcare Management*, 11(4), 356-362.
- Cochran, W.G. (1961). Comparison of methods for determining stratum boundaries. *Bulletin of the International Statistical Institute*, 35, 345-358.
- Cochran, W.G. (1977). *Sampling Techniques*. New York: John Wiley & Sons, Inc.
- Dalenius, T., et Gurney, M. (1951). The problem of optimum stratification. ii. *Scandinavian Actuarial Journal*, 1951(1-2), 133-148.
- Dalenius, T., et Hodges Jr., J.L. (1959). Minimum variance stratification. *Journal of the American Statistical Association*, 54(285), 88-101.

- Dayal, S. (1985). Allocation of sample using values of auxiliary characteristic. *Journal of Statistical Planning and Inference*, 11(3), 321-328.
- Edwards, D., Gilliland, D., Ward-Besser, G. et Lasecki, J. (2015). Conservative penny sampling. *Journal of Survey Statistics and Methodology*, 3(4), 504-523.
- Edwards, D., Ward-Besser, G., Lasecki, J., Parker, B., Wieduwilt, K., Wu, F. et Moorhead, P. (2003). The minimum sum method: A distribution-free sampling procedure for medicare fraud investigations. *Health Services and Outcomes Research Methodology*, 4(4), 241-263.
- Ekman, G. (1959). An approximation useful in univariate stratification. *The Annals of Mathematical Statistics*, 30(1), 219-229.
- Gilliland, D., et Feng, W. (2010). An adaptation of the minimum sum method. *Health Services and Outcomes Research Methodology*, 10(3), 154-164.
- Gunning, P., et Horgan, J.M. (2004). [Un nouvel algorithme pour la construction de bornes de stratification dans les populations asymétriques](https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2004002/article/7749-fra.pdf). *Techniques d'enquête*, 30, 2, 177-185. Article accessible à l'adresse <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2004002/article/7749-fra.pdf>.
- Gunning, P., Horgan, J.M. et Yancey, W. (2004). Geometric stratification of accounting data. *Contaduría y Administración*, (214), 0.
- Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *The Annals of Mathematical Statistics*, 1491-1523.
- Hedlin, D. (2000). A procedure for stratification by an extended ekman rule. *Journal of Official Statistics*, 16(1), 15.
- Hidiroglou, M.A. (1994). Sampling and estimation for establishment surveys: Stumbling blocks and progress. *Proceedings of the Survey Research Methods Section*, American Statistical Association, 693-698.
- Hidiroglou, M.A., et Kozak, M. (2018). Stratification of skewed populations: A comparison of optimisation-based versus approximate methods. *Revue Internationale de Statistique*, 86(1), 87-105.
- Hidiroglou, M.A., et Srinath, K. (1993). Problems associated with designing subannual business surveys. *Journal of Business & Economic Statistics*, 11(4), 397-405.

- Horgan, J.M. (2006). Stratification of skewed populations: A review. *Revue Internationale de Statistique*, 74(1), 67-76.
- Huffman, M. (2021). Value-based care: An executive briefing. *Nurse Leader*, 19(1), 82-86.
- Jurina, I., et Gligorova, L. (2017). Determination of the optimal stratum boundaries in the monthly retail trade survey in the croatian bureau of statistics. *Romanian Statistical Review*, (4).
- King, B. (1996). Sampling design issues when dealing with zeros. *Proceedings of the Survey Research Methods Section*, 400-405.
- King, B., et Madansky, A. (2013). On sampling design issues when dealing with zeros. *Journal of Survey Statistics and Methodology*, 1(2), 144-170.
- Kozak, M. (2004). Optimal stratification using random search method in agricultural surveys. *Statistics in Transition*, 6(5), 797-806.
- Kozak, M., et Verma, M.R. (2006). [Approche de la stratification par une méthode géométrique et par optimisation : Une comparaison de l'efficacité](https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2006002/article/9550-fra.pdf). *Techniques d'enquête*, 32, 2, 177-183. Article accessible à l'adresse <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2006002/article/9550-fra.pdf>.
- Lavallée, P., et Hidirolou, M.A. (1988). [Sur la stratification de populations asymétriques](https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/1988001/article/14602-fra.pdf). *Techniques d'enquête*, 14, 1, 35-45. Article accessible à l'adresse <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/1988001/article/14602-fra.pdf>.
- Li, X., et Ding, P. (2017). General forms of finite population central limit theorems with applications to causal inference. *Journal of the American Statistical Association*, 112(520), 1759-1769.
- Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97(4), 558-625.
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienne, Autriche. URL <http://www.R-project.org/>.
- Rao, J., et Wu, C. (2009). Empirical likelihood methods. Elsevier, *Handbook of Statistics*, 29, 189-207.
- Reddy, K.G., et Khan, M.G. (2019). Optimal stratification in stratified designs using weibull-distributed auxiliary information. *Communications in Statistics-Theory and Methods*, 48(12), 3136-3152.

Rivest, L. (1999). Stratum jumpers: Can we avoid them. *Proceedings of the Survey Research Methods Section*, American Statistical Association, 64-72.

Rivest, L.-P. (2002). [Une généralisation de l'algorithme de Lavallée et Hidioglou pour la stratification dans les enquêtes auprès des entreprises](https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2002002/article/6432-fra.pdf). *Techniques d'enquête*, 28, 2, 207-214. Article accessible à l'adresse <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2002002/article/6432-fra.pdf>.

Salmond, S.W., et Echevarria, M. (2017). Healthcare transformation and changing roles for nursing. *Orthopedic Nursing*, 36(1), 12.

Satterthwaite, F.E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2(6), 110-114.

Serfling, R. (1968). Approximately optimal stratification. *Journal of the American Statistical Association*, 63(324), 1298-1309.

Sethi, V. (1963). A note on optimum stratification of populations for estimating the population means. *Australian Journal of Statistics*, 5(1), 20-33.

Sigman, R.S., et Monsour, N.J. (1995). Selecting samples from list frames of businesses. *Business Survey Methods*, 133-152.

Singh, R. (1971). Approximately optimum stratification on the auxiliary variable. *Journal of the American Statistical Association*, 66(336), 829-833.

Sweet, E.M., et Sigman, R.S. (1995). Evaluation of model-assisted procedures for stratifying skewed populations using auxiliary data. *Proceedings of the Section on Survey Research Methods*, 1, 491-496.

Tschuprow, A.A. (1923). On the mathematical expectation of the moments of frequency distributions in the case of correlated observations. *Metron*, 2(461-493), 646-683.

Wang, M., et Aggarwal, V. (1984). Stratification under a particular pareto distribution. *Communications in Statistics-Theory and Methods*, 13(6), 711-735.