

## Article

# Fondements statistiques des enquêtes par téléphone mobile

par Kirk M. Wolter, Phil Smith et Stephen J. Blumberg

Décembre 2010



# Fondements statistiques des enquêtes par téléphone mobile

Kirk M. Wolter, Phil Smith et Stephen J. Blumberg<sup>1</sup>

## Résumé

Aux États-Unis, la taille de la population dotée d'un téléphone mobile seulement a augmenté rapidement ces dernières années et, par conséquent, les chercheurs ont commencé à expérimenter l'échantillonnage et l'interview des abonnés à la téléphonie mobile. Nous discutons des problèmes statistiques que posent les étapes de l'établissement du plan d'échantillonnage et de l'estimation des études par téléphone mobile. Les travaux sont exposés principalement dans le contexte d'une enquête à deux bases de sondage non chevauchantes dans laquelle une base et un échantillon sont employés pour la population possédant un téléphone fixe et une deuxième base de sondage et un deuxième échantillon, pour la population possédant seulement un téléphone mobile. Nous discutons également des aspects supplémentaires dont il faut tenir compte dans le cas d'une enquête à deux bases de sondage chevauchantes (où la base de sondage et l'échantillon pour la téléphonie mobile comprennent certains membres de la population dotée d'un téléphone fixe). Pour illustrer les méthodes, nous utilisons le plan de sondage de la National Immunization Survey (NIS) conçue pour surveiller les taux de vaccination chez les enfants de 19 à 35 mois et les adolescents de 13 à 17 ans. La NIS est une enquête téléphonique nationale, suivie d'une vérification des dossiers des fournisseurs de service, réalisée par les Centers for Disease Control and Prevention.

Mots clés : Étude par téléphone mobile ; composition aléatoire ; enquête à base de sondage double ; échantillonnage en réseau ; échantillonnage indirect ; règles de lien ; pondération des données d'enquête ; National Immunization Survey.

## 1. Introduction

Aux États-Unis, le nombre de personnes possédant un téléphone mobile a augmenté rapidement ces dernières années et le pourcentage d'adultes vivant dans un ménage équipé de téléphones mobiles devrait rapidement surpasser celui des adultes vivant dans un ménage muni de téléphones fixes (CTIA 2008 ; Blumberg et Luke 2008 ; Arthur 2007 ; Ehlen et Ehlen 2007). Par conséquent, les spécialistes de la recherche sur les sondages ont commencé à expérimenter des méthodes d'échantillonnage et d'interview des abonnés à la téléphonie mobile (Lavrakas, Shuttles, Steeh et Fienberg 2007). Le présent article porte sur les problèmes de conception et d'estimation statistique que posent les enquêtes par téléphone mobile. Y sont mises en relief des solutions rigoureuses sur le plan théorique, mais pratiques aux nouveaux problèmes que doivent résoudre aujourd'hui les spécialistes de la recherche sur les sondages dans le cadre des enquêtes par téléphone mobile.

Les enquêtes par téléphone conventionnelles, s'appuyant sur l'échantillonnage par composition aléatoire (CA), ne couvrent que la population de ménages munis d'au moins une ligne téléphonique fixe en état de fonctionnement utilisée pour les communications vocales. Dans une enquête par CA, l'hypothèse est que le téléphone fixe est un appareil électroménager et que chaque personne faisant partie de la population n'est attachée qu'à un seul ménage. Donc, les personnes peuvent être échantillonnées indirectement en

échantillonnant les numéros de téléphone et en partant de là pour appliquer des méthodes raisonnablement normalisées et bien connues d'estimation.

L'enquête par téléphone mobile entraîne un changement de paradigme et pose de nouveaux défis. Dans l'esprit de la plupart des gens, un téléphone mobile est un appareil personnel et non un appareil ménager. Certaines personnes partagent un téléphone mobile, dont 10 % à 20 % des adultes munis seulement d'un téléphone mobile (Carley-Baxter, Peytchev et Lynberg 2008), mais bon nombre ne le font pas et l'on ne peut donc pas supposer que tous les membres d'un ménage peuvent être rejoints au moyen de la même ligne téléphonique mobile. En outre, certains membres d'un ménage peuvent être rejoints au moyen de plus d'une ligne de téléphone mobile. Enfin, certains membres du ménage peuvent être rejoints uniquement au moyen d'une ligne de téléphone mobile, tandis que d'autres peuvent l'être au moyen d'un téléphone mobile et d'un téléphone fixe. Donc, dans les enquêtes par téléphone mobile, le ménage ne représente plus forcément la même organisation unifiante que dans le cas des enquêtes par téléphone conventionnelle.

Afin de lutter contre le risque croissant de biais (dû au sous-dénombrement) dans les enquêtes téléphoniques, on peut considérer des plans d'enquête téléphonique à base de sondage double qui comprennent un échantillon sélectionné par CA de lignes de téléphone fixes et un échantillon de lignes de téléphone mobiles. Les numéros de téléphone

1. Kirk M. Wolter, NORC et University of Chicago. Courriel : wolter-kirk@norc.org ; Phil Smith, National Center for Immunization and Respiratory Diseases ; Stephen J. Blumberg, National Center for Health Statistics.

figurant dans les deux bases de sondage sont non chevauchants, tandis que les personnes et les ménages correspondants qui peuvent être les sujets de l'enquête sont partiellement chevauchants.

Une théorie rigoureuse de l'estimation pour ce genre de plan de sondage téléphonique fait défaut, quoique certaines descriptions initiales de la pondération aient été avancées par Brick, Dipko, Presser, Tucker et Yuan (2006), Brick, Edwards et Lee (2007), et Frankel, Battaglia, Link et Mokdad (2007). Dans le présent article, nous présentons une théorie générale de l'estimation sans biais des totaux de population dans le contexte de plans d'enquête téléphonique à base de sondage double et calculons les poids de sondage correspondants. Nous montrons que l'information doit être recueillie durant l'enquête proprement dite pour permettre le calcul des poids d'échantillonnage.

Afin de fixer les idées, posons que  $A$  désigne la partie de la population globale d'intérêt accessible par la voie de la base de sondage des lignes de téléphone fixes,  $B$  désigne la partie accessible par la voie de la base de sondage des lignes de téléphone mobiles et  $C$  désigne la partie qui n'est accessible par la voie d'aucune des deux bases de sondage (la *population sans téléphone* et d'autres composantes relativement petites de la population totale). Soit  $a$  la sous-population dans  $A$  qui n'est pas accessible par la voie des lignes de téléphone mobiles (la *population à téléphone fixe seulement*), soit  $b$  la sous-population dans  $B$  non accessible par la voie des lignes téléphoniques fixes (la *population à téléphone mobile seulement*), et soit  $ab$  la sous-population accessible à la fois par ligne fixe et par ligne mobile (la *population à accès mixte*). Nous peaufinerons cette notation dans les sections qui suivent.

Savoir si une unité de la population d'intérêt est accessible par des lignes fixes ou des lignes mobiles est en soi une question complexe. Tout au long du présent article, quand nous disons qu'une unité est accessible par ligne fixe, nous entendons à la fois qu'il existe un accès à une ou à plusieurs lignes fixes (habituellement des lignes terrestres résidentielles uniquement) et qu'une personne répondrait effectivement à l'appel sur le téléphone fixe s'il sonnait pour des communications vocales. De nos jours, de nombreux adultes gardent une ligne téléphonique fixe strictement pour les communications informatiques et utilisent un téléphone mobile pour toutes les communications vocales. En vertu de notre définition, ces adultes ne sont pas considérés comme ayant accès à une ligne fixe et sont au contraire considérés comme faisant partie de la population à téléphone mobile seulement. De même, quand nous disons qu'une unité est accessible par ligne mobile, nous entendons qu'il existe à la fois l'accès physique à un téléphone mobile et l'intention de répondre à ce téléphone mobile s'il sonne. Toutes les autres unités de la population d'intérêt qui ne sont pas accessibles

par ligne fixe ou par ligne mobile sont considérées comme n'ayant pas de téléphone. Selon les données courantes, bien que personne n'en soit certain, de 20 % à 30 % environ des adultes font partie du domaine  $b$ , de 5 % à 10 %, du domaine  $C$ , et les autres sont répartis entre les domaines  $a$  et  $ab$ .

Ce que nous ont appris jusqu'à présent nos enquêtes par téléphone mobile et celles réalisées par d'autres est que la collecte des données est relativement coûteuse, le nombre moyen d'heures d'intervieweur par cas achevé étant de l'ordre de trois fois la moyenne observée pour les enquêtes conventionnelles par CA. Le coût plus élevé est dû, en partie, à l'exigence légale (aux États-Unis, la *Telephone Consumer Protection Act*) de composer manuellement les numéros de téléphone mobile sélectionnés. Les taux de réponse sont un peu plus faibles que ceux obtenus dans les enquêtes par CA. La durée de l'interview peut poser problème, certains répondants étant moins disposés à répondre à une longue interview par téléphone mobile que par téléphone fixe. En outre, des questions de protection de la vie privée peuvent constituer une contrainte dans le cas de l'interview par téléphone mobile, si la personne ne se trouve pas dans un lieu privé au moment de l'interview. La propension des utilisateurs de téléphone mobile à répondre peut varier de manière monotone avec leur niveau d'utilisation du téléphone mobile, les gros utilisateurs étant plus disposés à répondre que les utilisateurs légers ou occasionnels. La plupart des ruptures ont lieu durant les premières secondes de la tentative d'interview. Comme les enquêtes par téléphone mobile sont relativement nouvelles, les appels des intervieweurs sont inhabituels et ceux-ci n'ont que quelques secondes pour vendre l'enquête. Par ailleurs, nous constatons que de nombreux répondants rejoints sur téléphone mobile sont relativement prêts à coopérer une fois que l'on a réussi à capter leur attention en leur lisant l'introduction à l'enquête.

Étant donné toutes ces circonstances, nous considérons à l'heure actuelle l'échantillon de lignes mobiles comme un échantillon complémentaire relativement petit, l'échantillon principal continuant d'être un échantillon par CA plus grand de lignes fixes. L'échantillon de lignes mobiles est destiné à parachever la couverture de la population d'intérêt. Dans l'avenir, à mesure que le contexte des enquêtes par téléphone mobile évoluera vers la maturité et que les coûts baisseront, il sera peut-être possible de passer à une approche plus équilibrée où les échantillons de lignes fixes et de lignes mobiles seront de tailles similaires, voire même une situation où l'échantillon de lignes mobiles commencera à devenir dominant et où l'échantillon de lignes fixes sera utilisé comme complément pour achever la couverture.

À la section 2, nous décrivons les *réseaux d'unités d'échantillonnage*, d'*unités répondantes* et d'*unités d'estimation*, et montrons que les enquêtes par téléphone mobile

équivalent à un échantillonnage en réseau. À la section 3, nous présentons divers concepts dont nous aurons besoin durant la discussion de l'estimation d'après les données de sondage, entre autres la notion d'un *lien* (ou arête) entre les *nœuds* (ou sommets) du réseau. À la section 4, nous décrivons la dualité qui existe entre les populations correspondant aux différents types de nœuds. Notre approche rappellera à certains lecteurs les méthodes d'échantillonnage indirectes de Lavallée (2007). Le cœur de l'article est la section 5, où nous présentons les estimateurs sans biais des totaux de population pour les enquêtes par téléphone mobile et pour les plans d'enquête téléphonique à base de sondage double correspondants. À la section 6, à titre d'exemple, nous illustrons les incidences des nouvelles méthodes d'estimation dans le cas d'une enquête téléphonique existante sur la couverture de la vaccination des jeunes enfants et des adolescents. Nous concluons, à la section 7, par un bref résumé.

Tout au long de l'article, nous insistons sur l'élaboration de procédures d'échantillonnage et d'estimation rigoureuses, mais pratiques, pour la population *B*. Les méthodes des enquêtes par CA, c'est-à-dire les méthodes appliquées à la population *A*, sont bien connues et, dans une certaine mesure, ont été utilisées pendant des décennies ; pour une revue récente de ces méthodes, voir Wolter, Chowdhury et Kelly (2008).

## 2. Réseaux d'unités et le protocole de réponse

En général, au moins trois types d'unités existent dans le contexte d'une enquête par téléphone mobile, à savoir :

- les unités d'échantillonnage (UÉ) ;
- les unités répondantes (UR) ;
- les unités d'estimation (UE).

L'UÉ est l'unité d'échantillonnage de l'enquête. En pratique, les numéros de téléphone peuvent être échantillonnés directement d'après des listes de numéros de téléphone mobile ou l'être par étape, les indicatifs de central ou les banques de numéros servant d'unités primaires d'échantillonnage, et les numéros proprement dits étant sélectionnés en une ou plusieurs étapes de sous-échantillonnage dans les unités primaires. Pour simplifier la discussion, dans le présent article, nous considérons le numéro de téléphone proprement dit comme étant l'UÉ.

La cible réelle de l'interview de l'enquête, c'est-à-dire l'unité d'analyse, est ce que nous appelons l'unité d'estimation (UE). Certaines enquêtes sont axées sur la collecte et l'analyse de données sur les ménages ou sur les familles, le ménage ou la famille étant alors l'UE. D'autres visent à recueillir des données au niveau de la personne, auquel cas les personnes admissibles peuvent être des enfants de moins

de 18 ans, des adultes de 18 ans et plus, ou un segment démographique particulier de la population, tel que les femmes hispaniques de 0 à 34 ans. D'autres enquêtes encore sont axées sur les données au niveau du ménage ainsi qu'au niveau de la personne, et elles comportent dans ce cas au moins deux types d'UE et deux niveaux d'analyse.

Dans les enquêtes téléphoniques, l'adulte est l'unité répondante ou UR. L'UE peut ou non avoir la capacité de fournir directement les réponses à son sujet. Dans la négative, une UR répond en son nom. Si l'UE est un adulte, celui-ci, voire même un autre adulte peut servir d'UR correspondante. Si l'UE est un ménage, une famille, une unité de consommation ou un enfant, un ou plusieurs adultes peuvent servir d'UR correspondante. Le protocole de réponse, qui est spécifié par le méthodologiste de l'enquête, détermine effectivement quelles UR ont le droit de répondre pour quelles UE. Dans une enquête téléphonique habituelle, pour chaque UÉ sélectionnée dans l'échantillon, on communique par téléphone et on interviewe un répondant adulte (ou UR).

Dans une enquête par téléphone mobile, les liens entre les UÉ, les UR et les UE peuvent être différents. La figure 1 présente neuf réseaux qui illustrent certains types de liens possibles. Dans le premier réseau, une UÉ est liée à une UR, qui à son tour répond pour une UE. Cet arrangement pourrait avoir lieu si un adulte utilise une ligne téléphonique et que cet adulte à son tour fait la déclaration pour le ménage, ou pour lui-même, ou pour un enfant. Dans le deuxième réseau, une UÉ est liée à deux UR, qui peuvent répondre chacune pour l'UE. Cet arrangement s'observerait, par exemple, si deux adultes partageaient la même ligne téléphonique et que chacun aurait le droit, en vertu du protocole de l'enquête, de répondre au nom du ménage. Le cinquième réseau pourrait avoir lieu si deux adultes possédaient chacun leur propre ligne téléphonique non partagée avec l'autre, mais que chaque adulte de la paire avait le droit, en vertu du protocole de l'enquête, de répondre pour chacun des deux enfants.

Des réseaux plus compliqués sont possibles et doivent certainement exister de par le monde. Par exemple, le huitième réseau représente un arrangement où trois adultes partagent deux lignes téléphoniques. La première des lignes est partagée par les trois adultes, tandis que la deuxième est utilisée uniquement par le troisième adulte. Le premier des adultes a le droit, en vertu du protocole de l'enquête, de répondre pour deux UE, telles que ses enfants biologiques ; le deuxième n'a le droit de répondre pour aucune UE, et le troisième a le droit de répondre uniquement pour une troisième UE pour laquelle les deux premiers adultes ne peuvent pas faire de déclaration.

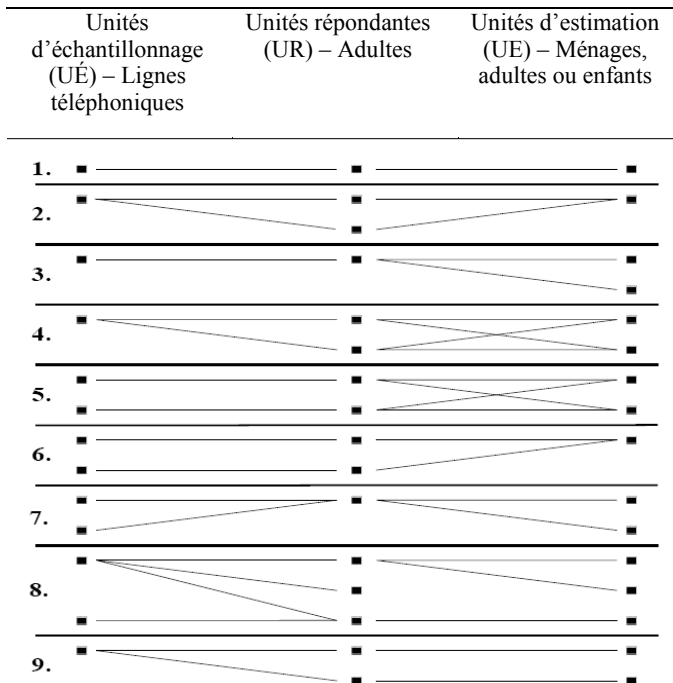


Figure 1 Exemples de réseaux dans une enquête par téléphone mobile

### 3. Liens entre les unités dans le réseau

Un *lien* est une relation saillante entre deux nœuds du réseau. Dans le contexte de la figure 1, les liens sont représentés par les segments de droite qui unissent les différents nœuds. Afin de fournir un fondement pour l'estimation d'après les données d'enquête, nous devons explorer les liens entre i) les UR et les UÉ, ii) les UE et les UR, et iii) les UE et les UÉ.

#### 3.1 Lien entre l'UR et l'UÉ

Deux concepts sont essentiels à la création d'un lien entre une UR et une UÉ, à savoir les concepts a) de *numéro de téléphone mobile personnel actif* (NMPA) et b) d'*accès habituel* à la ligne mobile.

Un NMPA correspond à une ligne téléphonique qui est en service au moment de l'enquête par téléphone mobile et dont les appels parviennent à un adulte admissible qui utilise le téléphone mobile, au moins partiellement, pour des raisons personnelles. Autrement dit, un NMPA doit satisfaire à trois critères :

- il est en service ;
- il permet d'entrer en communication avec un répondant adulte admissible ;
- il n'est pas utilisé exclusivement pour affaires.

Nous disons qu'un adulte donné a un accès habituel à un NMPA si, et uniquement si, cette personne possède l'usage

- régulier ;
- important ;
- continu de la ligne de téléphone mobile.

Chaque NMPA possède un ou plusieurs utilisateurs adultes réguliers et chaque utilisateur possède l'accès habituel à un ou à plusieurs téléphones mobiles. Dans de nombreux cas, il existe une relation un à un unique entre la ligne de téléphonie mobile et l'utilisateur adulte. Dans certains cas, par contre, il existe une relation un à plusieurs entre la ligne de téléphone mobile et ses utilisateurs.

Nous traitons une UÉ donnée et une UR donnée comme étant liées si, et uniquement si, l'UÉ est un NMPA et que l'UR a habituellement accès à l'UÉ. Une enquête par téléphone mobile doit tenir compte des liens qui existent entre la population d'UÉ et la population d'UR et reconnaître ces liens.

#### 3.2 Lien entre l'UE et l'UR

Une UE donnée est liée à une ou à plusieurs UR par les relations naturelles qui existent dans la société, telles que celles créées par la famille ou le lieu de résidence. Par exemple, un répondant adulte peut répondre à l'interview de l'enquête au nom de son ménage, de sa famille ou d'une unité de consommation. Il peut répondre en son propre nom, pour un enfant à charge de moins de 18 ans ou pour ses parents ou ses frères et sœurs.

Toute enquête nécessite un protocole de réponse qui précise quels répondants adultes doivent répondre pour quelles UE. Le protocole est choisi par le méthodologiste de l'enquête en tenant compte de la faisabilité, du coût et de l'exactitude de la déclaration. C'est ce protocole qui établit les liens entre les UE et les UR.

#### 3.3 Lien entre l'UE et l'UÉ

Les liens qui précèdent entre les UR et les UÉ, ainsi qu'entre les UE et les UR déterminent les liens entre les UE et les UÉ. Nous disons qu'une UE donnée est liée à une UÉ donnée si, et uniquement si, l'UE est liée à au moins une UR qui, à son tour, est liée à l'UÉ.

La notation qui suit sera utile dans l'exposé des sections suivantes. Soit  $j$  une UE donnée dans la population d'intérêt et soit  $i$  une UÉ donnée dans la population. Alors, définissons les variables indicatrices ou variables de lien

$$\ell_{ij} = \begin{cases} 1, & \text{si la } j^{\text{e}} \text{ UE est liée à la } i^{\text{e}} \text{ UÉ} \\ 0, & \text{autrement.} \end{cases}$$

#### 4. Dualité entre les populations d'UÉ et d'UE

Afin d'entreprendre l'élaboration d'une méthode d'estimation sans biais pour les enquêtes par téléphone mobile, nous établissons qu'une dualité existe entre la population d'UÉ, ou téléphones mobiles, (donc désignées par  $U^{\text{ÉB}}$ ) et la population d'UE qui sont liées aux téléphones mobiles (désignées par  $U^{\text{EB}}$ ). L'objectif d'une enquête par téléphone mobile est de faire des inférences concernant  $U^{\text{EB}}$ , mais nous verrons bientôt que cet objectif équivaut à faire certaines inférences concernant  $U^{\text{ÉB}}$  (dans cette notation, la première lettre de l'indice supérieur désigne le type d'unité, tandis que la lettre  $B$  désigne la base de sondage des lignes de téléphone mobiles. Plus loin, nous utiliserons l'indice  $A$  pour désigner la base de sondage des lignes de téléphone fixes).

Dans le domaine des UE, un total de population d'intérêt est donné par

$$Y^{\text{EB}} = \sum_{j \in U^{\text{EB}}} Y_j,$$

où, dans le deuxième membre, la variable  $Y$  est un item du questionnaire ou une autre variable, recodée ou dérivée, attachée aux unités de la population  $U^{\text{EB}}$ . De même, dans le domaine des UÉ, un total de population est défini par

$$X^{\text{ÉB}} = \sum_{i \in U^{\text{ÉB}}} X_i,$$

où, dans le deuxième membre, la variable  $X$  représente toute caractéristique fixe attachée aux unités de la population  $U^{\text{ÉB}}$ .

Alors que l'analyste des données de l'enquête s'intéresse au total pour la population d'UE (et à d'autres paramètres de cette population), nous pouvons obtenir un paramètre correspondant dans le domaine des UÉ en écrivant

$$Y^{\text{EB}} = \sum_{j \in U^{\text{EB}}} Y_j = \sum_{j \in U^{\text{EB}}} \sum_{i \in U^{\text{ÉB}}} \frac{Y_j \ell_{ij}}{\sum_{i' \in U^{\text{ÉB}}} \ell_{i'j}} = \sum_{i \in U^{\text{ÉB}}} X_i = X^{\text{ÉB}}, \quad (1)$$

où la variable  $X$  est maintenant définie spécifiquement par

$$X_i = \sum_{j \in U^{\text{EB}}} \frac{Y_j \ell_{ij}}{\sum_{i' \in U^{\text{ÉB}}} \ell_{i'j}}. \quad (2)$$

L'expression (1) montre la correspondance entre l'estimation dans le domaine des UÉ et l'estimation dans le domaine des UE. Le total  $X^{\text{ÉB}}$ , avec  $X_i$  définie comme en (2), est équivalent au total d'intérêt  $Y^{\text{EB}}$ , d'où le problème de l'estimation de  $Y^{\text{EB}}$  est équivalent au problème de l'estimation de  $X^{\text{ÉB}}$ .

Nous notons que (2) se présente essentiellement sous la même forme dans la théorie de l'échantillonnage indirect. Voir Lavallée (2007), Théorème 4.1. En échantillonnage

indirect, les UÉ sont liées à des grappes naturellement définies d'UE; si une UÉ donnée est sélectionnée dans l'échantillon, les données d'enquête sont recueillies pour toutes les UE comprises dans les grappes qui y sont liées. Ici, l'analogie est que les grappes sont définies par les UR qui répondent à la tentative d'interview par téléphone mobile et les données d'enquête sont recueillies auprès du répondant pour toutes les UE auxquelles il est lié. Ici, la situation est telle que la grappe est définie par la paire UÉ-UR. À cet égard, il se pose un problème d'identifiabilité qui n'existe généralement pas en échantillonnage indirect et nous examinons cette question plus en détail à la section 5.5.

Dans (2), nous attribuons effectivement une part égale de la variable  $Y_j$  à chaque UÉ  $i$  à laquelle la variable est liée. Nous pourrions aussi aboutir au même résultat en répartissant  $Y_j$  entre les UÉ auxquelles elle est liée, proportionnellement à une autre mesure connue de l'intensité de la relation entre  $j$  et  $i$ . Nous pourrions certes concevoir une répartition optimale de  $Y_j$  entre les UÉ qui sont liées, comme dans Deville et Lavallée (2006), mais ce genre de répartition peut être difficile à exécuter ou pourrait ne pas avoir beaucoup de poids dans des conditions pratiques à grande échelle.

#### 5. Estimation

Comme nous l'avons mentionné dans l'introduction, certaines UE seront liées exclusivement à des lignes de téléphone mobiles, certaines seront liées exclusivement à des lignes de téléphone fixes et certaines seront liées à la fois à des lignes de téléphone fixes et mobiles. Les UE sans téléphone, si tant est qu'il y en ait, ne seront liées ni à des lignes de téléphone mobiles ni à des lignes fixes. Pour établir la notation dans ce contexte, posons que  $U^{\text{E}}$  est la population globale d'UE d'intérêt et que  $U^{\text{E}}$  est la population globale d'UÉ. Soit  $U^{\text{EA}}$  les éléments de  $U^{\text{E}}$  qui sont liés à des lignes fixes, soit  $U^{\text{EB}}$  les éléments qui sont liés à des lignes mobiles, soit  $U^{\text{Ea}}$  les éléments qui sont liés uniquement à des lignes fixes, soit  $U^{\text{Eb}}$  les éléments qui sont liés uniquement à des lignes mobiles, soit  $U^{\text{Eab}}$  les éléments qui sont liés à la fois à des lignes fixes et à des lignes mobiles, et soit  $U^{\text{EC}}$  les éléments n'ayant pas le téléphone. Notons que  $U^{\text{E}} = U^{\text{EA}} \cup U^{\text{EB}} \cup U^{\text{EC}}$ ,  $U^{\text{EA}} = U^{\text{Ea}} \cup U^{\text{Eab}}$ , et  $U^{\text{EB}} = U^{\text{Eb}} \cup U^{\text{Eab}}$ , où  $U^{\text{Ea}}$ ,  $U^{\text{Eab}}$  et  $U^{\text{Eb}}$  sont des ensembles disjoints. En outre, soit  $U^{\text{ÉA}}$  la population de lignes de téléphone fixes, telles que  $U^{\text{E}} = U^{\text{ÉA}} \cup U^{\text{ÉB}}$ . Les lignes fixes et les lignes mobiles reflètent des sous-ensembles disjoints de la population globale d'UÉ.

Aux sections 5.1 et 5.2 qui suivent, nous discutons de l'estimation sans biais pour la sous-population, disons  $U^{\text{ET}} = U^{\text{EA}} \cup U^{\text{EB}}$ , qui est liée à au moins un téléphone de n'importe quelle sorte. Nous utilisons l'indice supérieur  $T$

pour désigner cette sous-population ayant le téléphone. Subséquentement, à la section 5.4, nous discutons brièvement de la couverture de la population sans téléphone.

Pour les UE comprises dans  $U^E$ , définissons les variables indicatrices

$$\begin{aligned} \delta_j &= 1, \text{ si aucune des UR liées à } j \text{ n'a accès à} \\ &\text{un service de téléphone fixe, tandis qu'au} \\ &\text{moins une des UR a habituellement accès} \\ &\text{à un service de téléphone mobile;} \\ &= 0, \text{ autrement.} \end{aligned}$$

$$\begin{aligned} \phi_j &= 1, \text{ si aucune des UR liées à } j \text{ n'a habituellement} \\ &\text{accès à un service de téléphone mobile,} \\ &\text{tandis qu'au moins l'une des UR a accès à} \\ &\text{un service de téléphone fixe;} \\ &= 0, \text{ autrement.} \end{aligned}$$

La variable  $\delta$  est indicatrice de la situation de possession d'un téléphone mobile seulement et la variable  $\phi$  est indicatrice de la situation de possession d'un téléphone fixe seulement.

Alors, le total de population d'intérêt peut être décomposé en

$$Y^{ET} = Y^{EA} + Y^{Eb}, \quad (3)$$

où

$$Y^{Eb} = \sum_{j \in U^{ET}} \delta_j Y_j$$

est le total pour le domaine des unités à téléphone mobile seulement, et

$$Y^{EA} = \sum_{j \in U^{ET}} (1 - \delta_j) Y_j$$

est le total du complément de ce domaine, qui comprend les UE liées exclusivement à des lignes fixes et les UE mixtes, qui sont liées à la fois à des lignes de téléphone fixes et mobiles. Le total des UE peut également s'écrire

$$Y^{ET} = Y^{Ea} + Y^{Eab} + Y^{Eb}, \quad (4)$$

où

$$Y^{Ea} = \sum_{j \in U^{ET}} \phi_j Y_j$$

est le total pour la population d'unités à téléphone fixe seulement, et

$$Y^{Eab} = \sum_{j \in U^{ET}} (1 - \delta_j) (1 - \phi_j) Y_j$$

est le total pour la population mixte ayant accès à une combinaison de téléphones fixes et de téléphones mobiles. Enfin, le total de population peut s'écrire

$$Y^{ET} = Y^{Ea} + Y^{Eb}, \quad (5)$$

où

$$Y^{Eb} = \sum_{j \in U^{ET}} (1 - \phi_j) Y_j$$

est le total pour le complément (dans la population ayant le téléphone) de la population ayant un téléphone fixe seulement.

Nous considérons (3) et, dans une certaine mesure, (4) comme les décompositions de l'intérêt pour les enquêtes téléphoniques et de l'importance de ces enquêtes, dans la pratique courante, aux États-Unis, et dans la suite de l'exposé, nous présentons des méthodes d'estimation de l'un et de l'autre. À l'heure actuelle, étant donné le coût relativement élevé des interviews par téléphone mobile, les enquêtes basées sur la décomposition (5) ne seraient pas rentables. Il serait presque toujours préférable de représenter le domaine  $U^{Eab}$  en utilisant un échantillon de lignes fixes plutôt qu'en utilisant un échantillon de téléphones mobiles. Si le coût relatif de l'interview par téléphone mobile tend à baisser dans l'avenir, la décomposition (5) pourrait devenir économiquement viable. Elle pourrait également l'être pour les enquêtes réalisées dans d'autres pays où la structure des coûts est plus favorable aux interviews par téléphone mobile.

### 5.1 Cas des domaines non chevauchants

À la présente section, nous utilisons un échantillon de lignes mobiles en vue de produire une estimation pour la population ayant un téléphone mobile seulement  $U^{Eb}$  et un échantillon de lignes fixes pour la production d'estimations pour la population complète de lignes de téléphone fixes  $U^{EA}$ . Nous observons qu'il est impossible de sélectionner directement un échantillon d'unités ayant une ligne de téléphone mobile seulement, parce que l'information sur la situation d'unité à téléphone mobile seulement n'est pas disponible dans la base de sondage, mais est plutôt recueillie au moyen du questionnaire de présélection de l'enquête. Pour appliquer ce plan, on éliminerait par présélection les répondants ayant un téléphone mobile qui se classent eux-mêmes dans le domaine mixte et on terminerait l'interview, en poursuivant uniquement avec les répondants possédant un téléphone mobile seulement.

Soit  $s^{EB}$  un échantillon probabiliste d'UÉ (lignes de téléphone mobiles) sélectionné dans la population  $U^{EB}$ , et soit  $\{W_i^{EB}\}$  l'ensemble de poids d'échantillonnage de base, tels que

$$\hat{X}^{EB} = \sum_{i \in s^{EB}} W_i^{EB} X_i$$

est un estimateur sans biais du total de population  $X^{EB}$ , où  $X_i$  est une caractéristique de la  $i^e$  unité de la population.

En émettant l'hypothèse d'un échantillonnage aléatoire simple sans remise dans les strates, les poids de base sont de la forme

$$W_i^{EB} = N_h / n_h, \quad (6)$$

où  $h$  désigne la strate d'échantillonnage dans laquelle la  $i^e$  UE est sélectionnée,  $N_h$  est le nombre d'UE présentes dans la base de sondage qui sont comprises dans la strate  $h$ , et  $n_h$  est la taille de l'échantillon tiré dans la strate  $h$ . Habituellement, la base de sondage des lignes de téléphone mobiles comprendrait tous les numéros de téléphone couverts par les indicatifs de central attribués par le système téléphonique aux téléphones mobiles. L'échantillonnage aléatoire simple serait la méthode la plus courante de sélection de l'échantillon pour ce genre d'indicatif. Peu d'information est disponible dans la base de sondage des lignes de téléphone mobiles pour permettre la stratification de l'échantillon, à part l'information géographique de niveau peu détaillé intégrée dans l'indicatif régional.

Soit  $s^{EB}$  l'échantillon correspondant d'UE, c'est-à-dire  $s^{EB} = \{j \in U^{EB} \mid j \text{ est liée à au moins une UE } i \text{ dans } s^{EB}\}$ . Nous utiliserons cet échantillon pour estimer le total de domaine des UE qui sont liées uniquement à un téléphone mobile,  $Y^{Eb}$ . Partant de (1) et (2), il est facile de voir que l'estimateur sans biais du total de domaine est donné par

$$\begin{aligned} \hat{Y}^{Eb} &= \sum_{i \in s^{EB}} W_i^{EB} \left\{ \sum_{j \in U^{EB}} \delta_j Y_j \ell_{ij} / \sum_{i' \in U^{EB}} \ell_{i'j} \right\} \\ &= \sum_{j \in s^{EB}} \delta_j Y_j W_j^{EB}, \end{aligned} \quad (7)$$

où les poids d'échantillonnage au niveau des UE sont définis par

$$W_j^{EB} = \sum_{i \in s^{EB}} W_i^{EB} \ell_{ij} / \sum_{i' \in U^{EB}} \ell_{i'j} \text{ pour } j \in s^{EB}. \quad (8)$$

De nouveau, voir Lavallée (2007) pour l'expression de ces poids dans le contexte de l'échantillonnage indirect.

Avant de quitter le domaine  $b$ , nous observons en passant qu'il est possible de sous-échantillonner les UE et de recueillir l'information d'enquête uniquement pour le sous-échantillon au lieu de recenser toutes les UE liées aux UR de l'échantillon. Si le statisticien choisissait une certaine forme de sous-échantillonnage, peut-être pour contrôler la taille d'échantillon ou le coût, un facteur de pondération supplémentaire apparaîtrait dans les poids donnés par (8). Ce genre de sous-échantillonnage est dénommé échantillonnage indirect à deux degrés dans Lavallée (2007, section 5.1).

Tournons-nous vers le domaine  $A$ . Soit  $s^{EA}$  un échantillon par CA standard de lignes de téléphone fixes, soit  $s^{EA}$  l'échantillon connexe d'UE, c'est-à-dire  $s^{EA} = \{j \in U^{EA} \mid j \text{ est liée à au moins une UE } i \text{ dans } s^{EA}\}$ , et soit

$$\hat{Y}^{EA} = \sum_{j \in s^{EA}} W_j^{EA} Y_j \quad (9)$$

l'estimateur sans biais standard du total de population. Par souci de concision, nous ne dériverons pas les poids d'échantillonnage standard ici ; pour plus de renseignements au sujet de ces poids, consulter Wolter et coll. (2008).

De (7) et (9) il découle que l'estimateur sans biais du total de population des UE est donné par

$$\hat{Y}^{ET} = \hat{Y}^{EA} + \hat{Y}^{Eb} \quad (10)$$

et les poids nécessaires pour appuyer cet estimateur sont  $\{W_j^{EA}\}$  et  $\{W_j^{EB}\}$ .

## 5.2 Cas des domaines chevauchants

Procédons maintenant à l'estimation en partant de la décomposition (4). Cela signifie que, dans l'échantillon de lignes de téléphone mobiles, nous interviewerons non seulement la population ne possédant qu'un téléphone mobile, mais aussi la population mixte (c'est-à-dire les personnes qui utilisent à la fois un téléphone fixe et un téléphone mobile). L'estimateur du total de population d'intérêt est maintenant de la forme

$$\hat{Y}^{ET} = \hat{Y}^{Ea} + \hat{Y}^{Eab} + \hat{Y}^{Eb}, \quad (11)$$

où

$$\hat{Y}^{Ea} = \sum_{j \in s^{EA}} W_j^{EA} \phi_j Y_j$$

est l'estimateur pour le domaine des unités à téléphone fixe seulement dérivées de l'échantillon de lignes de téléphone fixes,  $\hat{Y}^{Eb}$  est défini en (7) et est l'estimateur du domaine des unités à téléphone mobile seulement dérivées de l'échantillon de lignes de téléphone mobiles, et  $\hat{Y}^{Eab}$  est un estimateur du domaine mixte obtenu à partir des deux échantillons. L'estimateur pour le domaine mixte est

$$\begin{aligned} \hat{Y}^{Eab} &= \lambda \sum_{j \in s^{EA}} W_j^{EA} (1 - \phi_j) Y_j \\ &+ (1 - \lambda) \sum_{j \in s^{EB}} W_j^{EB} (1 - \delta_j) Y_j. \end{aligned} \quad (12)$$

Les poids nécessaires pour appuyer l'estimateur (11) sont  $\{W_j^{EA}\}$  et  $\{W_j^{EB}\}$ .

Voir Hartley (1962) pour une discussion du paramètre de mélange  $\lambda$  dans une enquête à base de sondage double, en se concentrant sur des considérations de variabilité d'échantillonnage. Si nous nous préoccupons maintenant du biais, Brick et coll. (2006) signalent que la propension à répondre à une enquête par téléphone mobile pourrait être liée positivement à la fréquence d'utilisation du téléphone mobile. Donc, les deux éléments du deuxième membre de



(12) pourraient être sujets à un biais de non-réponse différentiel qui n'est pas éliminé par la méthode classique des classes de pondération. Dans la population mixte, les utilisateurs peu fréquents du téléphone mobile pourraient être moins susceptibles de répondre s'ils sont interviewés dans l'échantillon de lignes mobiles que s'ils le sont dans l'échantillon de lignes fixes. Si ces adultes diffèrent considérablement des autres adultes dans la population mixte en ce qui concerne les caractéristiques clés étudiées dans l'enquête, l'estimateur (12), ainsi que l'estimateur (11) pourraient présenter un biais de non-réponse.

### 5.3 Estimation de la variance

Afin de faire des inférences concernant la population globale d'après l'échantillon, nous avons besoin d'un estimateur de la variance du total estimé. Premièrement, considérons le cas de domaines non chevauchants. En travaillant dans la population d'UÉ, nous pouvons employer des méthodes d'estimation de la variance appropriées pour le plan de sondage. Partant de (7), le total estimé pour le domaine des unités à téléphone mobile seulement peut s'écrire

$$\hat{Y}^{Eb} = \sum_{i \in S^{EB}} W_i^{EB} X_i,$$

où

$$X_i = \sum_{j \in U^{EB}} \delta_j Y_j \ell_{ij} / \sum_{i' \in U^{EB}} \ell_{ij}. \quad (13)$$

En supposant un échantillonnage aléatoire simple, l'estimateur sans biais de la variance du total estimé est donné par

$$v(\hat{Y}^{Eb}) = \sum_{h=1}^L N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{1}{n_h} s_{xh}^2,$$

où

$$s_{xh}^2 = \frac{1}{n_h - 1} \sum_{i \in S_h^{EB}} \left( X_i - \frac{1}{n_h} \sum_{i' \in S_h^{EB}} X_{i'} \right)^2.$$

Si nous ignorons le facteur de correction pour population finie, ce qui serait possible dans presque toute enquête téléphonique réelle, l'estimateur de variance devient

$$v(\hat{Y}^{Eb}) = \sum_{h=1}^L \frac{n_h}{n_h - 1} \sum_{i \in S_h^{EB}} \left( W_i^{EB} X_i - \frac{1}{n_h} \sum_{i' \in S_h^{EB}} W_{i'}^{EB} X_{i'} \right)^2. \quad (14)$$

Posons maintenant que  $v(\hat{Y}^{EA})$  est un estimateur de la variance de  $\hat{Y}^{EA}$  pour l'échantillon par CA de lignes fixes. Les estimateurs de ce genre étant bien connus, nous ne les passons pas en revue ici ; voir, par exemple, Wolter et coll.

(2008). Comme l'échantillonnage est effectué indépendamment dans les bases de sondage de lignes fixes et de lignes mobiles, l'estimateur sans biais de la variance du total estimé pour la population complète ayant le téléphone devient

$$v(\hat{Y}^{ET}) = v(\hat{Y}^{EA}) + v(\hat{Y}^{Eb}). \quad (15)$$

Afin de faciliter les développements qui suivent, soit  $\hat{V}^{EB}[\delta Y]$  un autre symbole pour représenter l'estimateur de variance donné par (14). Cette notation mettra l'accent sur le fait que l'estimateur de variance est basé sur la variable  $X_i$  donnée par (13) définie en fonction de la caractéristique  $\delta_j Y_j$ , qui est la caractéristique d'intérêt pour les UE ayant un téléphone mobile seulement. En outre, posons que le symbole  $\hat{V}^{EA}[Y]$  représente l'estimateur  $v(\hat{Y}^{EA})$  défini en fonction de la caractéristique  $Y_j$ . En nous servant de cette notation, (15) devient  $v(\hat{Y}^{ET}) = \hat{V}^{EA}[Y] + \hat{V}^{EB}[\delta Y]$ .

Deuxièmement, considérons l'estimation de la variance dans le cas de domaines chevauchants. L'estimateur du total pour la population ayant le téléphone est maintenant  $\hat{Y}^{ET}$  donné par (11). Pour une valeur fixe de  $\lambda$ , l'estimateur sans biais de la variance se dégage clairement du travail effectué dans (14) et (15). Il est donné par

$$v(\hat{Y}^{ET}) = \hat{V}^{EA}[\phi Y + \lambda(1 - \phi) Y] + \hat{V}^{EB}[\delta Y + (1 - \lambda)(1 - \delta) Y]. \quad (16)$$

Le premier terme du deuxième membre de (16) est l'estimateur de variance pour l'échantillon par CA de lignes fixes appliqué à la caractéristique composite  $\phi_j Y_j + \lambda(1 - \phi_j) Y_j$ , qui est la caractéristique pour les UE à téléphone fixe seulement plus une partie égale à  $\lambda$  de la caractéristique pour les UE mixtes. Le deuxième terme de (16) est l'estimateur de variance pour l'échantillon de lignes de téléphone mobiles appliqué à la caractéristique composite  $\delta_j Y_j + (1 - \lambda)(1 - \delta_j) Y_j$ , qui est la caractéristique pour les UE à téléphone mobile seulement plus une partie égale à  $(1 - \lambda)$  de la caractéristique pour les UE mixtes.

Les estimateurs des matrices de covariance peuvent être construits à partir d'expressions telles que (15) et (16), ce qui facilite l'inférence statistique concernant d'autres paramètres de population d'intérêt.

### 5.4 Correction des poids d'échantillonnage

Les poids d'échantillonnage peuvent être corrigés pour tenir compte de la non-réponse ou d'un calage planifié sur des totaux de contrôle connus.

Jusqu'à présent, nous ne nous sommes pas préoccupés des divers types de données manquantes qui peuvent survenir dans une enquête par téléphone mobile. Nous nous concentrerons ici sur le calcul des corrections pour les

données manquantes qui surviennent durant les interviews par téléphone mobile, en supposant que les corrections standard pour la présence de données manquantes dans l'échantillon de lignes de téléphone fixes ont déjà été intégrées dans les poids  $\{W_j^{EA}\}$ .

Les données manquantes peuvent être dues à trois facteurs, à savoir i) la *non-résolution* de l'UE ; ii) une *interview de présélection incomplète* de l'UR et iii) une *interview principale incomplète* de l'UR. Dans le présent article, nous adoptons les conventions suivantes. L'étape de résolution fait référence à la classification de l'UE comme un NMPA ou autre chose, par exemple une ligne déconnectée ou une ligne réservée aux affaires. Les UE dont la situation n'est pas résolue et les UE classées comme n'étant pas un NMPA ne poursuivent pas l'interview. L'étape de la présélection fait référence à une brève interview préliminaire destinée à confirmer la situation concernant le téléphone et à déterminer toute caractéristique démographique ou autre caractéristique d'admissibilité de toute UE liée à l'UR ; les UR pour lesquelles l'interview de présélection est incomplète ou pour lesquelles elle est complète, mais qu'aucune UE admissible n'est liée à l'UR ne passent pas à l'étape de l'interview. Si le protocole d'enquête requiert de n'inclure que les UE possédant un téléphone mobile seulement, comme à la section 5.1, l'interview se terminera à ce stade-là pour toute UE mixte. Par ailleurs, si le protocole d'enquête requiert d'inclure à la fois les UE à téléphone mobile seulement et les UE mixtes, comme à la section 5.2, alors l'interview se poursuivra pour toutes les UE de ce type. L'étape de l'interview fait référence à la collecte des réponses aux principaux items de l'enquête pour chaque UE admissible liée à l'UR. Le méthodologiste de l'enquête doit définir ce qui constitue une interview complète. En particulier, il doit décider si une *rupture* (une tentative d'interview qui est achevée pour certaines UE admissibles liées à l'UR, mais non toutes) doit être traitée comme une interview complète ou non. D'autres auteurs pourraient choisir d'organiser les étapes du processus de réponse à l'enquête d'une manière un peu différente de celle adoptée ici.

Les corrections des poids d'échantillonnage peuvent être faites pour la non-résolution et pour la non-réponse au questionnaire de présélection en supposant comme mécanisme de réponse que les données manquent au hasard. Ces deux corrections doivent être faites au niveau de l'UE. Soit  $\{s_\alpha^{EB}\}$  une partition de l'échantillon de lignes de téléphone mobiles en cellules de pondération  $\alpha$  spécifiées par l'utilisateur et désignons maintenant les poids d'échantillonnage tirés de (6) par  $W_{1i}^{EB}$ , où l'indice inférieur 1 a été ajouté simplement pour indiquer la première étape d'un processus de correction en plusieurs étapes. Les indicatifs téléphoniques régionaux et les variables de recensement géographiques au niveau de l'indicatif régional peuvent être

utilisés pour former les cellules de pondération ; sinon, peu d'information sur les covariables est disponible en ce qui concerne les numéros de téléphone mobile. Les taux d'achèvement de la résolution particuliers aux téléphones mobiles sont définis par

$$R_{1\alpha} = \frac{\sum_{i \in s_\alpha^{EB}} r_{1i} W_{1i}^{EB}}{\sum_{i \in s_\alpha^{EB}} W_{1i}^{EB}},$$

où  $r_{1i}$  est une variable indicatrice de résolution (= 1, si le cas est résolu, = 0, s'il n'est pas résolu), et les poids corrigés de la non-résolution sont  $W_{2i}^{EB} = r_{1i} W_{1i}^{EB} / R_{1\alpha}$  pour  $i \in s_\alpha^{EB}$ .

Soit  $e_{1i}$  une variable indiquant si  $i$  est un NMPA résolu (= 1, si NMPA résolu, = 0, autrement), et soit  $\{s_\beta^{EB}\}_{\beta=1}^B$  une partition de l'échantillon de lignes de téléphone mobiles en cellules de pondération spécifiées par l'utilisateur, qui pourraient être les mêmes que celles de la partition qui précède, ou être différentes. Alors, les taux d'achèvement de l'interview de présélection particuliers aux téléphones mobiles sont donnés par

$$R_{2\beta} = \frac{\sum_{i \in s_\beta^{EB}} r_{2i} e_{1i} W_{2i}^{EB}}{\sum_{i \in s_\beta^{EB}} e_{1i} W_{2i}^{EB}},$$

où  $r_{2i}$  est une variable indicatrice de présélection (= 1, si la présélection est achevée, = 0, si la présélection n'est pas achevée), et les poids corrigés de la non-réponse à la présélection sont  $W_{3i}^{EB} = r_{2i} e_{1i} W_{2i}^{EB} / R_{2\beta}$  pour  $i \in s_\beta^{EB}$ . Notons que la somme appropriée des poids est préservée à chaque étape du processus de correction.

Ensuite, les poids d'échantillonnage doivent être corrigés de la non-réponse à l'interview. Selon la classification adoptée pour les ruptures d'interview par le méthodologiste de l'enquête, deux cas doivent être pris en considération : i) l'UR achève ou n'achève pas l'interview pour l'ensemble des UE qui sont admissibles et qui lui sont liées, ou ii) l'UR achève ou n'achève pas l'interview de manière sélective sur la base des UE individuelles. Si les ruptures sont classées comme des interviews incomplètes, seul le cas i s'applique. Soit  $e_{2i}$  une variable indiquant si l'UR est présélectionnée et si elle est liée à au moins une UE qui est admissible à l'interview (= 1, si présélectionnée et admissible, = 0, autrement), et soit  $r_{3i}$  la variable indicatrice d'interview (= 1, si l'interview est complète, = 0, autrement).

Pour le cas i, la correction des poids peut être faite au niveau de l'UE et est donnée par  $W_{4i}^{EB} = r_{3i} e_{2i} W_{3i}^{EB} / R_{3\gamma}$  pour  $i \in s_\gamma^{EB}$ , où  $R_{3\gamma}$  est le taux pondéré d'achèvement de l'interview calculé à l'intérieur des cellules de pondération  $\gamma$  spécifiées par l'utilisateur. De nouveau, les options pour la construction des cellules de pondération sont limitées

dans une enquête par téléphone mobile ; elles peuvent être spécifiées en fonction de l'information disponible à l'étape de pondération précédente ou de toute information recueillie durant l'interview de présélection. Le taux pondéré d'achèvement de l'interview est

$$R_{3\gamma} = \frac{\sum_{i' \in s_{\gamma}^{EB}} r_{3i'} e_{2i'} W_{3i'}^{EB}}{\sum_{i' \in s_{\gamma}^{EB}} e_{2i'} W_{3i'}^{EB}}.$$

Le total estimé pour le domaine des unités ayant un téléphone mobile seulement peut maintenant s'exprimer par

$$\hat{Y}^{Eb} = \sum_{j \in s^{EB}} \delta_j Y_j W_{4j}^{EB}, \tag{17}$$

où

$$W_{4j}^{EB} = \sum_{i \in s^{EB}} W_{4i}^{EB} \ell_{ij} / \sum_{i' \in U^{EB}} \ell_{i'j}$$

et  $s^{EB}$  est l'ensemble des UE admissibles déclarées durant les interviews de présélection. Le poids est nul pour toute UE admissible comprise dans  $s^{EB}$  pour laquelle l'UR n'a pas achevé de répondre à l'interview principale. Le total estimé pour le domaine des unités mixtes, s'il est requis en fonction du protocole d'enquête, est défini de la même façon par

$$\hat{Y}^{Eab} = \lambda \sum_{j \in s^{EA}} W_j^{EA} (1 - \phi_j) Y_j + (1 - \lambda) \sum_{j \in s^{EB}} W_{4j}^{EB} (1 - \delta_j) Y_j.$$

Pour le cas ii, la correction de la non-réponse à l'interview doit être faite au niveau de l'UE. Les UE sont traitées comme des cas engendrés et il faut décider pour chacune s'il s'agit d'un cas d'interview achevée. Pour le domaine des unités à téléphone mobile seulement, le total estimé est (17), où le poids est maintenant défini par

$$W_{4j}^{EB} = r_{3j} e_{2j} W_{3j}^{EB} / R_{3\gamma} \text{ pour } j \in s^{EB},$$

$$W_{3j}^{EB} = \sum_{i \in s_3^{EB}} W_{3i}^{EB} \ell_{ij} / \sum_{i' \in U^{EB}} \ell_{i'j},$$

et

$$R_{3\gamma} = \frac{\sum_{j' \in s_{\gamma}^{EB}} r_{3j'} W_{3j'}^{EB}}{\sum_{j' \in s_{\gamma}^{EB}} W_{3j'}^{EB}}.$$

Ici, les cellules de pondération,  $\gamma$ , sont définies en fonction des caractéristiques des UE, déterminées d'après l'interview de présélection ou d'autres sources.

Aussi bien pour le cas i que ii, afin de faciliter les calculs, posons que  $W_{4j}^{EA}$  est défini et égal à 0 pour les UE comprises dans l'échantillon de lignes de téléphone mobiles et

que  $W_{4j}^{EB}$  est égal à 0 pour les UE comprises dans l'échantillon de lignes de téléphone fixes. Si le protocole d'enquête est celui de la section 5.1, nous concluons que les poids de sondage pour l'estimation du total de population d'intérêt sont définis par

$$W_j = W_{4j}^{EA} + W_{4j}^{EB} \delta_j \tag{18}$$

pour  $j \in s^{ET}$ , où  $s^{ET} \in s^{EA} \cup s^{EB}$ . Autrement, si le protocole d'enquête est celui de la section 5.2, nous concluons que les poids de sondage sont définis par

$$W_j = W_{4j}^{EA} \{\phi_j + \lambda(1 - \phi_j)\} + W_{4j}^{EB} \{\delta_j + (1 - \lambda)(1 - \delta_j)\} \tag{19}$$

pour  $j \in s^{ET}$ .

Les poids corrigés de la non-réponse donnés par (18) ou (19) peuvent être calés (Deville et Särndal 1992) sur des totaux de contrôle externes dans les cellules socioéconomiques ou géographiques pour la population d'UE, en appliquant des méthodes de poststratification, de calage (ratissage) ou d'estimation par la régression généralisée (GREG). Si des sources précises sont disponibles, les totaux de contrôle peuvent être établis et le calage peut être exécuté séparément pour les domaines  $A$  et  $b$  ou pour les domaines  $a$ ,  $ab$  et  $b$ . Si l'on ne dispose pas de totaux de contrôle selon la situation concernant le téléphone, le calage doit se faire en utilisant des totaux de contrôle pour la population complète, indépendamment de la situation concernant le téléphone.

Afin d'illustrer ces idées, nous examinons brièvement l'estimateur GREG. Supposons que nous disposions d'une variable auxiliaire  $1 \times p$ ,  $\mathbf{Z}_j$ , pour les UE admissibles observées pour lesquelles les totaux de contrôle  $\mathbf{Z}^{ET} = \sum_{j \in U^{ET}} \mathbf{Z}_j$  sont connus. Par exemple, la variable  $z$  peut être issue d'un modèle entièrement saturé en ce qui concerne les variables explicatives d'âge, de race et de sexe. Soit  $s_4^{ET}$  l'ensemble d'UE pour lesquelles l'interview principale est achevée et soit  $n_4^{ET} = \#(s_4^{ET})$  le nombre d'UE admissibles déclarées durant les interviews achevées obtenues dans l'échantillon consolidé de lignes téléphoniques. Empilons les valeurs de  $y$ , les valeurs de  $z$  et les poids dans les matrices  $\mathbf{Y} = (Y_1, \dots, Y_{n_4^{ET}})'$ ,  $\mathbf{Z} = (\mathbf{Z}'_1, \dots, \mathbf{Z}'_{n_4^{ET}})'$  et  $\mathbf{W} = \text{diag}(W_1, \dots, W_{n_4^{ET}})'$ . Alors, l'estimateur GREG (Cassel, Särndal et Wretman 1976) du total d'intérêt pour la population ayant le téléphone prend la forme familière

$$\tilde{Y}^{ET} = \hat{Y}^{ET} + (\mathbf{Z}^{ET} - \hat{\mathbf{Z}}^{ET}) \hat{\beta} = \sum_{j \in s_4^{ET}} W_j g_j Y_j,$$

où les coefficients estimés sont donnés par  $\hat{\beta} = (\mathbf{Z}'\mathbf{W}\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{W}\mathbf{Y}$ ,  $\hat{Y}^{ET} = \sum_{j \in s_4^{ET}} W_j Y_j$ ,  $\hat{\mathbf{Z}}^{ET} = \sum_{j \in s_4^{ET}} W_j \mathbf{Z}_j$  et  $g_j = 1 + (\mathbf{Z}^{ET} - \hat{\mathbf{Z}}^{ET}) \mathbf{Z}'_j$ . Lavallée (2007, chapitre 7) dérive l'estimateur par développement en série de Taylor de

la variance de l'estimateur GREG dans un contexte d'échantillonnage indirect. Voir aussi Wolter (2007, chapitre 6) pour l'estimation de la variance de l'estimateur GREG.

Avant d'abandonner le sujet du calage, notons que nous avons en grande partie laissé de côté la petite population n'ayant pas le téléphone, qu'il est fondamentalement impossible d'échantillonner dans une enquête téléphonique. Pourtant, selon toute vraisemblance, le total global de population  $Y^E = Y^{ET} + Y^{EC}$  sera le paramètre d'intérêt plutôt que le total de la population ayant le téléphone  $Y^{ET}$ , et les totaux de contrôle connus utilisés dans le calage pourraient être les totaux pour la population globale  $Z^E = Z^{ET} + Z^{EC}$ , et non les totaux pour la population ayant le téléphone  $Z^{ET}$ . Afin d'inclure la population sans téléphone, nous pourrions envisager d'utiliser un estimateur GREG révisé avec  $g_j = 1 + (Z^E - \hat{Z}^{ET})Z'_j$ . Cette révision s'appuie sur le même modèle pour la population sans téléphone que pour la population ayant le téléphone. Voir Keeter (1995) et Chowdhury, Montgomery et Smith (2008) pour d'autres considérations concernant le calage des poids pour la population sans téléphone.

### 5.5 Hypothèses d'identifiabilité

La théorie qui précède repose sur l'hypothèse fondamentale que, si l'UE  $i$  est sélectionnée dans l'échantillon de lignes de téléphone mobiles, alors la variable  $X_i$  définie en (2) est observable. Néanmoins, le 9<sup>e</sup> réseau (comme le 8<sup>e</sup>) de la figure 1 illustrent un problème éventuel de cette théorie. Dans le cas de ce réseau, deux UR sont liées à une UE et, à leur tour, les UR sont liées chacune à une seule UE. Pour poursuivre cet exemple, nous supposons que ces deux UE ne sont liées à aucune autre UR dans la population. Au moment de l'interview de l'enquête, une seule des UR sera habituellement rejointe et interviewée (à moins que le protocole d'enquête exige spécifiquement qu'une tentative d'interview ait lieu auprès de chaque UR liée à l'UE sélectionnée). L'UR répondante fournira les renseignements pour l'UE à laquelle elle est liée, mais étant donné la nature même de ce réseau, le répondant ne peut pas faire de déclaration pour l'UE qui est liée à l'UR compagnon qui partage la ligne de téléphone mobile figurant dans l'échantillon. Donc, il existe au moins une UE liée à l'UE qui ne peut pas être observée, c'est-à-dire pour laquelle les données ne peuvent pas être recueillies durant l'interview par téléphone mobile. Donc, nous disons que  $X_i$  est *non identifiable*. La situation concernant la possibilité de déclaration pour les deux UE serait inversée si, au moment de la tentative d'interview par téléphone mobile, l'appel était reçu par l'UR compagnon.

Afin de préserver l'absence de biais de l'estimateur du total de population, la variable  $X_i$  doit être identifiable pour chaque UE répondante sélectionnée dans l'échantillon de

lignes de téléphone mobiles. Nous devons formuler l'une de deux hypothèses. Premièrement, nous pourrions supposer que nous éliminons le problème en agissant comme si des réseaux tels que les réseaux 8 et 9 n'existent pas ou que leur nombre est négligeable.

Deuxièmement, une approche plus réaliste consisterait à émettre l'hypothèse d'une étape de randomisation supplémentaire, à savoir que la tentative d'appel pour l'interview devant être administrée à l'UE a atteint une UR liée à l'UE sélectionnée aléatoirement. Cette randomisation pourrait être considérée comme étant conceptuelle (autrement dit, survenant naturellement et non dirigée par le méthodologiste de l'enquête). De manière formelle et rigoureuse, nous devrions recueillir l'information sur le nombre d'UR liées à l'UE et la probabilité que la tentative d'appel sur le téléphone mobile soit reçue par l'UR répondante. La probabilité serait approximée par la part d'utilisation du téléphone mobile autodéclarée par le répondant. Si une UR seulement est liée à l'UE, cette probabilité est de 1,0 et, manifestement, cette valeur simple ne devra pas être recueillie durant l'interview une fois qu'il a été signalé qu'il n'existe qu'une UR. Si deux UR ou plus sont liées à l'UE, la probabilité ou la part dont la valeur doit être recueillie est désignée par  $\tau_{ik}$  pour les UR portant l'indice inférieur  $k$ , où  $\sum_{k \in U_i^{RB}} \tau_{ik} = 1$  et  $U_i^{RB}$  est l'ensemble d'UR qui sont liées à la  $i^e$  UE. En disposant de cette information supplémentaire, un estimateur sans biais de

$$X_i = \sum_{j \in U^{EB}} \frac{\delta_j Y_j \ell_{ij}}{\sum_{i' \in U^{EB}} \ell_{i'j}}$$

est donné par

$$\hat{X}_i = \sum_{j \in U^{EB}} \frac{1}{\sum_{i' \in U^{EB}} \ell_{i'j}} \sum_{k \in U_i^{RB}} \alpha_{ik} \frac{\delta_j Y_j \ell_{ij} \ell_{ikj}}{\tau_{ik} \sum_{k' \in U_i^{RB}} \ell_{ik'j}}, \quad (20)$$

où  $\alpha_{ik}$  est une variable indicatrice signalant si la  $k^e$  UR est le répondant réalisé ou non pour la  $i^e$  UE dans  $s^{EB}$  et

$$\begin{aligned} \ell_{ikj} &= 1, & \text{si l'UE } i \text{ est lié à l'UR } k \text{ qui, à son} \\ & & \text{tour, est liée à l'UE } j; \\ &= 0, & \text{autrement.} \end{aligned}$$

Les données sont maintenant identifiées et nous pouvons insérer (20) dans (7), ce qui donne l'estimateur révisé

$$\hat{Y}^{EB} = \sum_{j \in s^{EB}} \delta_j Y_j W_{0j}^{EB} \quad (21)$$

avec les poids révisés

$$W_{0j}^{EB} = \sum_{i \in s^{EB}} W_i^{EB} \frac{1}{\sum_{i' \in U^{EB}} \ell_{i'j}} \sum_{k \in U_i^{RB}} \alpha_{ik} \frac{\ell_{ij} \ell_{ikj}}{\tau_{ik} \sum_{k' \in U_i^{RB}} \ell_{ik'j}}. \quad (22)$$

En guise d'approximation, nous pourrions supposer que les UR sont égales aux utilisateurs de téléphone mobile,

auquel cas  $\tau_{ik}$  serait simplement la réciproque du nombre d'UR liées à l'UE  $i$  pour toutes les UR  $k$ . La correction de la non-réponse et le calage sur les totaux de contrôle seraient effectués comme auparavant.

Alternativement, le méthodologiste de l'enquête pourrait demander une étape de randomisation réelle, qui nécessiterait que l'intervieweur dresse la liste des UR liées à l'UE et qu'il sélectionne une de ces UR au hasard, ou une étape de pseudorandomisation en utilisant la méthode du dernier anniversaire. De telles méthodes ne sont probablement pas applicables en ce moment, à cause de la difficulté à obtenir la collaboration des personnes sélectionnées pour les interviews par téléphone mobile.

### 5.6 Incidences sur la collecte des données

Certaines données doivent être recueillies durant l'interview afin de permettre le calcul des estimateurs dont nous discutons ici.

Pour appuyer l'utilisation de  $\delta_j$ , l'enquête par téléphone mobile doit fournir des renseignements permettant d'établir si une des UR liées à l'UE a accès à un téléphone fixe. L'UR répondante doit fournir cette information en ce qui la concerne ainsi que pour d'autres UR qui pourraient être liées à l'UE.

Afin d'appuyer l'utilisation de  $\phi_j$ , l'enquête par ligne téléphonique fixe doit fournir des renseignements permettant d'établir si l'une des UR liées à l'UE possède un accès régulier à un téléphone mobile. L'UR répondante doit fournir cette information en ce qui la concerne ainsi que pour d'autres UR qui pourraient être liées à l'UE. Cette déclaration peut être assez simple si le protocole de réponse relie uniquement les UE aux UR présentes dans le même ménage. Pour des protocoles de réponse plus compliqués, la déclaration pourrait être difficile à obtenir.

Afin d'appuyer l'utilisation de  $\sum_{i \in U^{ib}} \ell_{ij}$  dans le calcul des poids de sondage, des données doivent être recueillies durant l'enquête afin d'établir combien d'UE dans la population sont liées à l'UE  $j$  pour laquelle est faite la déclaration. L'UR répondante doit être capable de donner le nombre de téléphones mobiles, y compris le sien, dont les appels parviennent à une UR qui est liée à une UE donnée.

Si l'estimateur donné par (21) et (22) est utilisé afin d'identifier toutes les UE, des renseignements supplémentaires doivent être recueillis durant l'interview. L'UR répondante doit connaître et indiquer le nombre d'UR, y compris elle-même, qui sont liées à la fois à l'UE sélectionnée et à l'UE pour laquelle est faite la déclaration. L'UR répondante doit également connaître et indiquer sa part de l'utilisation du téléphone mobile sur lequel l'interview est effectuée ou être capable de préciser que l'utilisation est approximativement égale.

## 6. Exemple : la National Immunization Survey (NIS)

Afin d'illustrer l'information qui doit être recueillie durant l'interview de l'enquête, nous nous servons de la NIS, qui est une enquête réalisée auprès des parents d'enfants de 19 à 35 mois et d'adolescents de 13 à 17 ans parrainée par les Centers for Disease Control and Prevention (CDC) afin de surveiller les taux de couverture de la vaccination (c'est-à-dire la proportion d'enfants dont la vaccination est à jour en regard du calendrier recommandé de vaccination) aux États-Unis. Les données recueillies dans le cadre de la NIS le sont en deux phases, à savoir une enquête téléphonique par CA auprès des ménages munis d'un téléphone fixe qui ont de jeunes enfants ou des adolescents dans la tranche d'âge admissible, suivie d'une enquête par la poste auprès des prestataires de la vaccination des enfants d'âge admissible. Pour la phase de l'enquête téléphonique de la NIS, la base de sondage comprend l'ensemble des numéros de téléphone au États-Unis dans les banques non vides de 100 numéros de téléphone consécutifs (banques de type 1+). À l'heure actuelle, les numéros de téléphone mobile regroupés sous les indicatifs de central réservés à la téléphonie mobile ne sont pas inclus dans la base d'échantillonnage de la NIS. Quand un ménage comptant un enfant d'âge admissible est repéré durant l'enquête téléphonique, l'interview est effectuée auprès de l'adulte dans le ménage qui est reconnu comme étant la personne la mieux informée au sujet de la situation de vaccination de l'enfant (presque toujours la mère ou le père). Durant l'interview téléphonique, des données sont recueillies pour chaque enfant d'âge admissible présent dans le ménage, y compris les caractéristiques démographiques de l'enfant, les caractéristiques démographiques de la mère et les caractéristiques socioéconomiques du ménage où vit l'enfant. À la fin de l'interview téléphonique, l'intervieweur demande l'autorisation de communiquer avec le prestataire de la vaccination de l'enfant. Si l'autorisation est accordée, tous les prestataires de vaccination nommés par le répondant durant l'interview téléphonique sont contactés par la poste afin d'obtenir les antécédents de vaccination de l'enfant déclarés par le prestataire, qui sont utilisés dans l'analyse statistique pour évaluer la situation de vaccination. Smith, Hoaglin, Battaglia, Khare et Barker (2005) décrivent en détail les méthodes statistiques utilisées par l'équipe de la NIS.

Étant donné la croissance de la population ayant un téléphone mobile seulement, la proportion de la population cible de la NIS qui est couverte par la base de sondage des lignes de téléphone fixes a diminué ces dernières années. En utilisant les données de la *National Health Interview Survey*, Khare, Singleton, Wouhib et Jain (2008) estiment

qu'environ 18 % des enfants admissibles et 10 % des adolescents admissibles pourraient ne pas figurer dans la base de sondage de la NIS. Pour faire face à l'accroissement du nombre de ménages dotés d'un téléphone mobile seulement dans la population cible de la NIS, des interviews par téléphone mobile pourraient être ajoutées à l'enquête.

Dans le cas de la NIS, le numéro de téléphone est l'UÉ, la mère ou le père bien informé est l'UR et l'enfant admissible est l'UE. Pour l'échantillon par CA ou échantillon A, le parent est un membre du ménage auquel le numéro de ligne fixe dans l'échantillon est attribué, tandis que pour l'échantillon de lignes de téléphone mobiles ou échantillon B, le parent possède l'accès régulier au téléphone mobile auquel le numéro de téléphone de l'échantillon est attribué. Dans la NIS, au lieu de sous-échantillonner les enfants, le parent le mieux informé fournit les renseignements pour tous les enfants d'âge admissible qui vivent dans son foyer (mais pas pour les enfants qui pourraient vivre ailleurs). Ces éléments du protocole d'enquête établissent les liens entre les UR et les UÉ et entre les UE et les UR.

Un plan de sondage exhaustif pour la NIS consiste à effectuer l'estimation en se basant sur les domaines non chevauchants et la décomposition (3). Autrement dit, l'échantillon A est utilisé pour représenter tous les enfants liés à un ménage possédant un téléphone fixe et l'échantillon B est utilisé pour représenter tous les enfants liés à un parent possédant un téléphone mobile seulement. Nous avons considéré et rejeté les décompositions (4) et (5) pour des raisons de coût et de risque de biais de non-réponse différentiel dans l'estimation pour la population mixte.

Pour appliquer l'estimateur donné par (10), nous déterminons si l'enfant compris dans l'échantillon A vit dans un ménage doté d'un téléphone fixe seulement en utilisant les trois questions suivantes :

- A1. Maintenant, je vais vous poser quelques questions au sujet des téléphones mobiles dans votre ménage. En tout, combien de téléphones mobiles en état de fonctionnement avez-vous, ainsi que les membres de votre ménage, à votre disposition pour usage personnel ? Veuillez ne pas compter les téléphones mobiles qui sont utilisés exclusivement pour affaires.
- A2. Combien de [ces] téléphones mobiles les parents et gardiens de [ÉNUMÉRER TOUS LES ENFANTS ADMISSIBLES] utilisent-ils habituellement ?
- A3. Les appels téléphoniques que vous et les membres de votre famille recevez, sont-ils presque tous reçus sur des téléphones mobiles, presque tous reçus sur des téléphones ordinaires, ou certains sont-ils reçus sur des téléphones mobiles et certains, sur des téléphones ordinaires ? (SI LE RÉPONDANT DEMANDE S'IL FAUT INCLURE LES APPELS

D'AFFAIRES : Veuillez ne pas inclure dans votre réponse les appels liés aux affaires).

Pour l'échantillon de lignes de téléphone mobiles ou échantillon B, nous déterminons si l'enfant fait partie d'un ménage à téléphone mobile seulement au moyen des deux questions suivantes.

- B1. Avez-vous une ligne de téléphone fixe dans votre ménage ? (INTERVIEWEUR : APPROFONDISSEZ SI OUI : Veuillez ne pas inclure les lignes réservées uniquement aux modems, les lignes réservées uniquement aux télécopieurs, les lignes utilisées uniquement pour un système de sécurité à domicile, les téléavertisseurs ou les téléphones mobiles).
- B2. En pensant uniquement au téléphone raccordé à la ligne fixe, et non à votre téléphone mobile, si ce téléphone sonnait et que quelqu'un se trouvait à la maison, dans les circonstances normales, quelle est la probabilité que l'on réponde à l'appel ? Diriez-vous très probable, assez probable, assez improbable ou tout à fait improbable ?

Nous utiliserions la question B2, due à Cantor, Brownlee, Zukin et Boyle (2008), pour déterminer si la ligne de téléphone fixe est effectivement utilisée pour des communications vocales et, donc, si le répondant appartient au domaine *ab* ou *b*.

Également pour l'échantillon B, afin de déterminer le nombre de téléphones mobiles dans la population qui sont liés à un enfant admissible donné, nous utiliserions les deux questions suivantes :

- B3. Maintenant, je vais vous poser quelques questions au sujet des téléphones mobiles dans votre ménage. En tout, combien de téléphones mobiles en état de fonctionnement avez-vous, ainsi que les membres de votre ménage, à votre disposition pour usage personnel ? Veuillez ne pas compter les téléphones mobiles qui sont utilisés exclusivement pour affaires, et veuillez inclure le numéro que nous avons appelé.
- B4. Combien de [ces] téléphones mobiles les parents et gardiens de [ÉNUMÉRER ENFANTS] utilisent-ils habituellement ? Veuillez inclure le numéro que nous avons appelé.

Les réponses aux questions A1 à A3 et B1 à B4 permettent de calculer les poids de sondage et d'appliquer l'estimateur sans biais du total de population donné par (10).

## 7. Résumé

Dans le présent article, nous avons utilisé certains éléments théoriques de l'échantillonnage indirect et de l'échantillonnage en réseau pour démontrer un cadre statistique pour la conception et l'analyse d'enquêtes par téléphone mobile. Nous avons présenté un estimateur sans biais du total de population en ce qui concerne les unités d'estimation liées aux unités d'échantillonnage. Par implication, cette théorie offre un moyen de construire des estimateurs pour d'autres paramètres de population qui peuvent être exprimés sous forme de fonction de totaux. Nous avons illustré les problèmes en prenant comme exemple la NIS, qui est une enquête téléphonique portant sur les jeunes enfants et les adolescents.

Des renseignements recueillis durant l'interview de l'enquête sont nécessaires pour classer les unités d'estimation dans le domaine des utilisateurs de téléphones mobiles seulement, le domaine des utilisateurs de téléphones fixes seulement ou le domaine mixte. L'erreur de déclaration pourrait donner lieu à des erreurs de classification et nuire à l'absence de biais de l'estimateur, comme le pourrait la non-réponse à l'enquête dans les interviews par téléphone mobile et par téléphone fixe.

## Remerciements

Les auteurs remercient le rédacteur en chef adjoint pour ses précieux commentaires.

## Bibliographie

- Arthur, A. (2007). The birth of a cellular nation. *The Source*. Mediamark Research Inc. Disponible au [http://www.mediamark.com/mri/TheSource/sorc2007\\_09.htm](http://www.mediamark.com/mri/TheSource/sorc2007_09.htm), 3.
- Blumberg, S.J., et Luke, J.W. (2008). Wireless substitution: Early release of estimates from the National Health Interview Survey. National Center for Health Statistics. Disponible au <http://www.cdc.gov/nchs/nhis.htm>.
- Brick, J.M., Dipko, S., Presser, S., Tucker, C. et Yuan, Y. (2006). Nonresponse bias in a dual frame sample of cell and landline numbers. *Public Opinion Quarterly*, 70, 780-793.
- Brick, J.M., Edwards, W.S. et Lee, S. (2007). Sampling telephone numbers and adults, interview length, and weighting in the California Health Interview Survey cell phone pilot study. *Public Opinion Quarterly*, 71, 793-813.
- Cantor, J., Brownlee, S., Zukin, C. et Boyle, J. (2008). Do We Need to Worry About Wireless Substitution in Public Opinion Polls about Health Reform. Présentation au AcademyHealth 25<sup>th</sup> Annual Research Meeting, Washington, DC.
- Carley-Baxter, L., Peytchev, A. et Lynberg, M. (2008). Comparison of cell phone and landline surveys: A design perspective. Document présenté au Annual meeting of the American Association for Public Opinion Research, New Orleans, LA.
- Cassel, C.-M., Särndal, C.-E. et Wretman, J.H. (1976). Some results on generalized difference estimation and generalized regression estimation for finite populations. *Biometrika*, 63, 615-620.
- Chowdhury, S., Montgomery, R. et Smith, P.J. (2008). Adjustment for noncoverage of nonlandline telephone households in and RDD Survey. *Proceedings of the Survey Research Methods Section*, American Statistical Association, Alexandria, VA.
- CTIA (2008). Wireless Quick Facts. Disponible au <http://www.ctia.org/advocacy/research/index.cfm/AID/10323>.
- Deville, J.-C., et Lavallée, P. (2006). Sondage indirect : les fondements de la méthode généralisée du partage des poids. *Techniques d'enquête*, 32, 185-196.
- Deville, J.-C., et Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376-382.
- Ehlen, J., et Ehlen, P. (2007). Cellular-only substitution in the United States as lifestyle adoption: Implications for telephone survey coverage. *Public Opinion Quarterly*, 71, 717-733.
- Frankel, M., Battaglia, M., Link, M. et Mokdad, A. (2007). Integrating cell phone numbers into Random Digit-Dialed (RDD) landline surveys. *Proceedings of the Survey Research Methods Section*, American Statistical Association, (Alexandria, VA), 3793-3800.
- Hartley, H.O. (1962). Multiple frame surveys. *Proceedings of the Social Statistics Section*, American Statistical Association, 203-206.
- Keeter, S. (1995). Estimating non-coverage bias from a phone survey. *Public Opinion Quarterly*, 59, 196-217.
- Khare, M., Singleton, J.A., Wouhib, A. et Jain, N. (2008). Assessment of Potential Bias in the National Immunization Survey (NIS) from the Increasing Prevalence of Households Without Landline Telephones. Présentation au National Immunization Conference, Centers for Disease Control and Prevention.
- Lavallée, P. (2007). *Indirect Sampling*. New York : Springer Science+Business Media, LLC.
- Lavrakas, P.J., Shuttles, C.D., Steeh, C. et Fienberg, H. (2007). The state of surveying cell phone numbers in the United States: 2007 and Beyond. *Public Opinion Quarterly*, 71, 840-854.
- Smith, P.J., Hoaglin, D.C., Battaglia, M.P., Khare, M. et Barker, L.E. (2005). Statistical methodology of the National Immunization Survey, 1994-2002. National Center for Health Statistics, Hyattsville, MD. *Vital and Health Statistics*, Série 2, 138.
- Wolter, K.M. (2007). *Introduction to Variance Estimation, Second Edition*. New York : Springer-Verlag.
- Wolter, K.M., Chowdhury, S. et Kelly, J. (2008). Design, conduct, and analysis of random digit dialing surveys. Dans *Handbook of Statistics: Sample Surveys, Theory, Methods and Inference*, (Éds., D. Pfeffermann et C.R. Rao), Elsevier, Oxford, UK.