

Article

Symposium 2008 :
Collecte des données : défis, réalisations et nouvelles orientations

La non-réponse dans le cadre d'une enquête à composition aléatoire: l'expérience du cycle 21 (2007) de l'Enquête sociale générale

par Isabelle Marchand, Ryan Chepita, Patrick St-Cyr et Kuawa Williams

2009



La non-réponse dans le cadre d'une enquête à composition aléatoire: l'expérience du cycle 21 (2007) de l'Enquête sociale générale

Isabelle Marchand, Ryan Chepita, Patrick St-Cyr et Kuawa Williams¹

Résumé

La difficulté grandissante à joindre les répondants affecte, de façon générale, la non-réponse des enquêtes téléphoniques, et particulièrement les enquêtes qui utilisent les techniques de composition aléatoire (CA) comme l'Enquête sociale générale (ESG). L'ESG est une enquête annuelle à sujets multiples effectuée auprès de 25 000 répondants. Ses objectifs sont de surveiller les caractéristiques et les modifications importantes de la structure sociale canadienne. Le cycle 21 (2007) de l'ESG porte sur la famille, le soutien social et la retraite et vise la population âgée de 45 ans et plus vivant dans les 10 provinces canadiennes. Pour cibler plus efficacement cette population, une partie de l'échantillon provient d'un suivi auprès des répondants du cycle 20 (2006) de l'ESG qui portait sur les transitions familiales. Il est complété par un nouvel échantillon obtenu par CA. Dans cet article, nous présentons d'abord le plan de sondage de l'enquête et la méthode de composition aléatoire utilisée. Nous décrivons ensuite les défis reliés au calcul du taux de non-réponse dans une enquête à composition aléatoire qui cible un sous-ensemble d'une population, pour lequel la population dans le champ de l'enquête doit être estimée ou modélisée. Ceci est fait notamment avec l'utilisation de paradonnées. La méthodologie utilisée au cycle 21 de l'ESG est présentée en détail.

Mots clés : Enquête à composition aléatoire, taux de réponse, cas non-résolus, sous-population.

1. Introduction

L'Enquête sociale générale (ESG) est une enquête annuelle transversale effectuée auprès de la population. Elle a été mise sur pied par Statistique Canada en 1985 pour combler certaines lacunes dans les statistiques sociales canadiennes. Afin de répondre à une gamme variée de préoccupations sociales et politiques du gouvernement fédéral, l'enquête a deux objectifs principaux: premièrement, rassembler des données pour suivre l'évolution de la structure sociale et du mode de vie des Canadiens; deuxièmement, fournir des renseignements immédiats sur des statistiques sociales précises. Pour ce faire, le programme de l'ESG consiste à mener une enquête touchant cinq grands thèmes principaux. Chaque thème est abordé aux cinq ans au moyen d'une enquête annuelle, appelée cycle. Ce délai est jugé raisonnable pour mesurer l'évolution, les changements sociaux s'effectuant sur une longue période de temps. Les derniers thèmes étaient la famille, les réseaux sociaux, l'emploi du temps, la victimisation et la retraite. Chaque questionnaire d'un cycle de l'ESG comporte trois modules: un contenu commun qui est répété aux cinq ans pour assurer le volet tendance (thème principal), un contenu spécifique (thème particulier) qui varie d'un cycle à l'autre pour assurer le volet immédiat et un dernier module de classification, présent à chaque cycle et qui se compose d'un ensemble de variables de base de données démographiques et de variables socio-économiques. À ses débuts, l'enquête était menée auprès de 10 000 personnes. Depuis le cycle 13 (1999) elle est effectuée auprès de 25 000 personnes. L'enquête utilise le système d'interviews téléphoniques assistées par ordinateur (ITAO), par le biais de la méthode de composition aléatoire de numéros de téléphone.

Le cycle 21 de l'ESG (ESG-21) porte sur la famille, le soutien social et la retraite, et cible la population âgée de 45 ans et plus vivant dans les dix provinces canadiennes. Pour rejoindre plus efficacement cette population, le plan de sondage utilise deux composantes. La première partie de l'échantillon provient d'un suivi des répondants de l'ESG-20 potentiellement éligibles à l'ESG-21; la seconde partie provient d'un nouvel échantillon obtenu à l'aide des techniques de CA.

¹ Isabelle Marchand, Ryan Chepita, Patrick St-Cyr et Kuawa Williams, Statistique Canada, Immeuble R.H. Coats 16^e étage, 100 promenade Tunney's Pasture, Ottawa Ontario, Canada, K1A 0T6

L'article est organisé comme suit: le plan d'échantillonnage de l'ESG-21 et la méthode de CA utilisée par l'enquête sont présentés dans les deux prochaines sections. La section 4 décrit le défi relié au calcul du taux de réponse dans le cadre d'une enquête à CA qui cible un sous-ensemble de la population. La section 5 expose en détail la méthodologie utilisée par l'ESG-21 pour le calcul du taux de réponse de la composante CA. La section 6 illustre la méthode pour le calcul du taux de réponse de la composante provenant de l'ESG-20 et du taux de réponse global de l'ESG-21.

2. Plan d'échantillonnage de l'ESG-21

La population cible de l'ESG-21 comprend toute personne âgée de 45 ans et plus vivant dans les 10 provinces canadiennes à l'exclusion des pensionnaires à temps plein des établissements. Parmi les exigences du plan de sondage de l'ESG, on retrouve celle de fournir des estimations détaillées au niveau national et celle d'assurer une comparabilité urbaine et rurale par province. Afin de répondre aux objectifs de l'enquête, la population canadienne est stratifiée par provinces, elles-mêmes divisées en régions métropolitaines de recensement (RMR), groupe de RMR ou non-RMR. Le Canada comprend 27 strates au total. À l'intérieur de chaque strate, la répartition de l'échantillon est effectuée de façon à permettre de balancer les besoins de données provinciales et nationales. Cette répartition est faite selon la méthode de Kish (Kish, 1976). Chaque taille initiale est ensuite gonflée pour tenir compte de la non-réponse et de l'éligibilité du ménage puisque seuls les ménages ayant au moins un membre de 45 ans et plus sont éligibles à l'ESG-21. La taille est finalement ajustée pour tenir compte du fait que l'unité d'échantillonnage est un numéro de téléphone, et non le ménage.

Tel que mentionné à la section 1, une partie de l'échantillon provient d'un suivi de l'ESG-20. Le cycle 20 portait sur les transitions familiales et visait la population âgée de 15 ans et plus vivant dans les 10 provinces canadiennes. Cet échantillon a également été sélectionné en utilisant les techniques de CA. Les données ont été recueillies au cours de l'année 2006 et un taux de réponse de 67,4 % a été obtenu. Parmi tous les répondants de l'ESG-20, 13 230 ont été identifiés potentiellement éligibles d'appartenir à l'ESG-21, c.-à-d. âgés de 43 ans et plus, et font partie de l'échantillon du cycle 21. On estime obtenir 10 500 répondants de cette façon. Le reste de l'échantillon, soit 14 500 répondants, provient d'un nouvel échantillon sélectionné par CA. La collecte de l'ESG-21 s'est effectuée de mars à décembre 2007. L'échantillon annuel est fractionné en neuf échantillons mensuels qui ont chacun une période de collecte qui s'étend sur une période de deux mois appelée vague. L'échantillon est ainsi distribué afin de représenter les variations saisonnières de l'information.

3. La composition aléatoire

3.1 Base de sondage et sélection de l'échantillon

L'ESG utilise l'approche par CA et en particulier, la technique d'élimination des banques invalides. Au Canada, les numéros de téléphone sont formés de 10 chiffres qui peuvent être décomposés en un code régional (trois premiers chiffres), un préfix (trois suivants) et un identificateur de banque (deux chiffres suivants). Une banque est donc formée des huit premiers chiffres et contient 100 numéros de téléphone possibles. Par exemple, (613) 951-4703 appartient à la banque 61395147. La base de sondage est formée d'une liste de banques valides. Une banque valide contient par définition au moins un numéro de téléphone résidentiel selon nos données administratives. À Statistique Canada (StatCan), les données de InfoDirect et les fichiers de facturation de différentes compagnies sont utilisés pour identifier les numéros de téléphone résidentiels et par conséquent, les banques valides. Aux fins d'échantillonnage, chaque banque est attribuée à une strate géographique.

La sélection de l'échantillon se fait à l'aide d'un tirage aléatoire simple avec remise de banques à l'intérieur de chaque strate. On génère ensuite de façon aléatoire les deux derniers chiffres pour obtenir un numéro de téléphone. Pour chaque numéro de téléphone sélectionné on lui attribue le statut « répertorié » s'il apparaît dans nos répertoires téléphoniques ou « inconnu » s'il n'y apparaît pas.

3.2 Caractéristiques

En ce qui a trait à la couverture de la base de sondage, elle exclue les ménages n'utilisant aucun service téléphonique. Selon l'Enquête sur le service téléphonique résidentiel (ESTR) 2007, cette proportion de la population est estimée à 0,9 %. Les ménages utilisant les téléphones cellulaires seulement sont également exclus. Cette proportion de la population est estimée à 6,4 %, toujours selon l'ESTR 2007. Étant donné que nous utilisons la méthode d'élimination des banques invalides, les numéros résidentiels appartenant aux banques faussement assignées invalides sont aussi exclus. Cette partie est jugée négligeable.

Au cours des dernières années, on observe une diminution des taux de réponse dans les enquêtes par interviews téléphoniques, en particulier celles qui utilisent l'approche par CA. Il est effectivement de plus en plus difficile de rejoindre les gens et de les convaincre de participer à une enquête par téléphone. C'est aussi une approche où on retrouve peu de variables auxiliaires pour optimiser le plan de sondage et traiter la non-réponse. La portabilité des numéros de téléphone diminue l'exactitude de l'information géographique et ceci peut à long terme contribuer à diminuer l'efficacité du plan de sondage. On retrouve par contre certains avantages de l'approche par CA: il est en effet relativement rapide de développer une enquête à CA et de la mettre en œuvre. C'est aussi une approche qui est moins coûteuse qu'une interview sur place. Du point de vue statistique, c'est une méthode qui donne lieu à de petits effets de plan.

3.3 Collecte

Parmi les numéros de téléphone sélectionnés dans l'échantillon, ceux qui ont un statut « inconnu » sont traités par une compagnie externe à StatCan pour subir un premier tri appelé processus de pré-collecte. À l'aide d'un système automatisé de pré-composition, les numéros de téléphone hors-service sont identifiés. Cette étape permet de nettoyer l'échantillon : environ 24 % de l'échantillon initial est enlevé suite à ce processus. Ensuite, les numéros de téléphone sont envoyés à la collecte régulière. Depuis 2006, pour gérer le fardeau de réponse, une limite d'appels est imposée aux enquêtes. Pour les enquêtes à CA, une limite de vingt appels est assignée aux numéros qui ont le statut « répertoire » et une limite de cinq appels est attribuée aux numéros de téléphone au statut « inconnu ». Il est à noter que si un contact est détecté à l'intérieur de ces cinq appels, la limite d'appels est réinitialisée à vingt.

Dans l'approche par composition aléatoire, une partie importante de l'échantillon comprend des numéros de téléphone qui ne rejoignent pas un ménage. On estime la proportion de numéros de téléphone envoyés sur le terrain qui rejoindra un ménage à 54 %. Pour le cycle 21, un ménage est éligible si au moins un de ses membres est âgé de 45 ans et plus. Cette proportion était estimée à 64 %. On retrouve donc pour l'ESG-21, un taux de succès global attendu de 35 % (=54 % x 64 %).

4. Taux de réponse

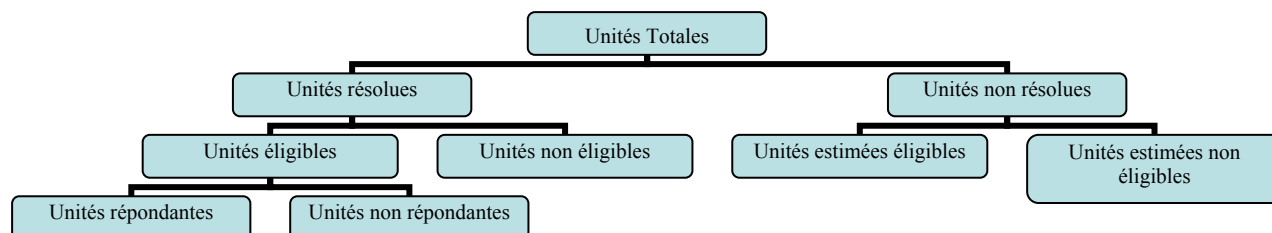
4.1 Calcul du taux de réponse

Un des principaux indicateurs pour évaluer la qualité d'une enquête est le taux de réponse. Celui-ci illustre la portion de l'enquête qui est touchée par la non-réponse. La non-réponse peut être une source importante d'erreur non échantillonnale si elle introduit un biais dans les estimations. En effet, si les répondants n'ont pas les mêmes caractéristiques que les non répondants par rapport à une variable d'intérêt, et que l'imputation ou la repondération, par modèle de non-réponse ou par calage, ne permet pas d'atténuer suffisamment la différence entre ces deux, les estimations ou les conclusions analytiques pourraient être erronées. Bien que le lien ne soit pas direct entre biais et faible taux de réponse, il demeure que le calcul du taux de réponse est un indicateur essentiel pour informer les utilisateurs et les gestionnaires d'enquête de la qualité des données.

La Politique visant à informer les utilisateurs sur la qualité des données et la méthodologie a pour but de fournir aux utilisateurs les renseignements requis pour évaluer la conformité des données d'enquête à leur utilisation (Statistique Canada, 2000). À cette fin, selon les normes et lignes directrices de Statistique Canada pour la déclaration des taux de non-réponse (Statistique Canada, 2001), les utilisateurs doivent être informés du taux de réponse dans le cadre de la phase d'estimation. Dans ce contexte, les unités non résolues suite au processus de collecte doivent être attribuées

éligibles ou non éligibles à l'enquête. La Figure 4-1 illustre la répartition des cas à l'étape de l'estimation des données.

Figure 4-1
Composantes répondants / non-répondants à l'étape de l'estimation des données



Le taux de réponse fourni aux utilisateurs est calculé selon l'équation suivante:

$$\text{Taux de réponse} = \frac{\text{Unités répondantes}}{\text{Unités éligibles} + \text{Unités estimées éligibles}}$$

4.2 Les unités non résolues de la composante CA de l'ESG-21

Le défi du calcul du taux de réponse dans une enquête à CA est la difficulté de statuer sur l'éligibilité des unités. Tel que mentionné à la section 3.3, dans une enquête CA, on s'attend à ce que seulement une partie des numéros de téléphone envoyés à la collecte joigne une résidence privée. Les autres numéros vont identifier des unités non éligibles ou seront classés comme unités non résolues. Les unités non éligibles sont les numéros de téléphone qui desservent des numéros hors-service, des entreprises, des téléphones cellulaires, etc. Les unités non résolues correspondent aux numéros de téléphone qui rejoignent des lignes occupées, des télécopieurs (qui peuvent potentiellement rejoindre une résidence privée), des numéros sans réponse, des répondants sans indication claire de résidence privée ainsi que tout autre problème technique.

Différents facteurs contribuent à une hausse du nombre d'unités non résolues. Par exemple, la mobilité, le peu de disponibilité et le peu de volonté des répondants affectent négativement la possibilité à établir le contact. Aussi, les progrès dans le domaine des télécommunications fait en sorte que des outils comme les répondeurs et les afficheurs permettent de filtrer les appels menant ainsi à l'évitement. Ils font aussi en sorte que des numéros de téléphone peuvent être utilisés à la fois pour les télécopieurs, Internet et les lignes résidentielles terrestres, augmentant ainsi la difficulté à statuer sur certains cas. De plus, la limite d'appels fixée fait en sorte que certains cas qui auraient pu être résolus en quelques appels additionnels se voient attribuer le statut de non résolu.

Pour la composante CA du cycle 21, puisqu'un sous-ensemble de la population générale est visé, la catégorie des unités non résolues comprend tous les numéros de téléphone qui ont permis de joindre une résidence, mais qui, pour une raison quelconque, n'ont pas rempli la matrice de composition du ménage. La matrice de composition du ménage est une série de questions qui recueillent des données sur l'âge, le sexe et la relation entre tous les membres du ménage. Pour être considéré comme une unité éligible, le ménage doit être composé d'au moins une personne de 45 ans et plus et les données de la matrice de composition sont essentielles pour confirmer l'éligibilité du ménage. Par exemple, une personne rejointe dans sa résidence privée qui refuse de répondre à l'ESG-21 avant de remplir la matrice de composition du ménage est une unité non résolue pour l'ESG-21. Parmi les 57 741 numéros de téléphone envoyés à la collecte, environ 18 % sont considérés comme non résolu. Ce volume important à traiter a nécessité le développement d'une méthodologie particulière.

5. Méthodologie du calcul du taux de réponse de la composante CA de l'ESG-21

5.1 Aperçu de la méthodologie générale

Pour l'ESG-21, la méthodologie retenue pour la résolution des cas s'apparente à un processus d'imputation, le statut étant attribué au niveau micro pour chaque unité échantillonnale considérée comme une unité non résolue. L'imputation se fait en 2 étapes. On assigne tout d'abord un statut d'éligibilité / non-éligibilité « résidence », c.-à-d. on détermine si le numéro de téléphone non résolu joint une résidence privée ou non. Ensuite, pour tous les numéros de téléphone, imputés ou non, joignant une résidence et pour lesquels la matrice de composition du ménage est incomplète ou n'est pas remplie, on assigne un statut d'éligibilité / non-éligibilité « âge », correspondant au fait que le ménage compte au moins une personne âgée de 45 ans ou plus. La méthodologie se décompose en deux parties: dans un premier temps, une série de règles déterministes permet de statuer directement sur l'éligibilité du cas. Dans un 2^e temps, une étape de modélisation assigne un statut d'éligibilité de façon aléatoire.

La principale source d'information utilisée pour résoudre les cas non résolus est les paradonnées. Les paradonnées sont les données issues du processus de collecte, qui en plus d'aider à la gestion de la collecte, aide à comprendre et éventuellement à optimiser le processus de collecte (Laflamme, 2008). Les paradonnées de l'ESG fournissent environ une quarantaine de codes de résultats possibles pour chaque tentative d'appels tels que: aucun contact, ligne occupée, télécopieur, refus, etc. De plus, un vecteur de longueur 1 à 20 est associé avec chaque numéro de téléphone envoyé à la collecte. Ce vecteur contient le code de résultat obtenu pour chaque tentative effectuée pour ce numéro de téléphone, ainsi que le statut final de collecte du cas : soit « final », « en-progrès » ou « limite d'appels atteinte ».

5.2 Règles déterministes

À cette étape, les règles déterministes permettent de statuer directement sur l'éligibilité des cas. Les sources d'information utilisées sont les paradonnées (les codes de résultats possibles et le statut du cas) et les données d'enquête. Les codes de résultat possibles ont été regroupés en trois catégories: 1-« résidence » c.-à-d. ceux qui déterminent que le numéro de téléphone est relié à une résidence (par ex: barrière linguistique, refus, rendez-vous, etc.), 2-« non résidence », c.-à-d. ceux qui déterminent que le numéro de téléphone a joint une unité qui n'est pas une résidence (par ex: téléphone cellulaire, logement collectif, entreprise) et 3-« incapable de statuer », c.-à-d. ceux qui ne permettent pas de conclure (par ex: aucun contact, répondeur, télécopieur). Pour ce qui est des données d'enquête, les données recueillies au début de l'interview peuvent permettre de confirmer qu'il y a eu un contact avec une résidence privée ou non, et les données provenant de la matrice de composition du ménage sont obligatoires pour assigner une éligibilité « âge ». Pour ce qui est de la construction des règles, une approche hiérarchique a été retenue où la priorité est accordée aux cas finalisés sur le terrain et aux codes de résultats assignés par les interviewers. Les tableaux suivants illustrent les règles pour chaque niveau d'éligibilité à déterminer.

Tableau 5.2-1**Résumé des règles déterministes utilisées pour la composante CA de l'ESG-21 - Éligibilité « résidence »**

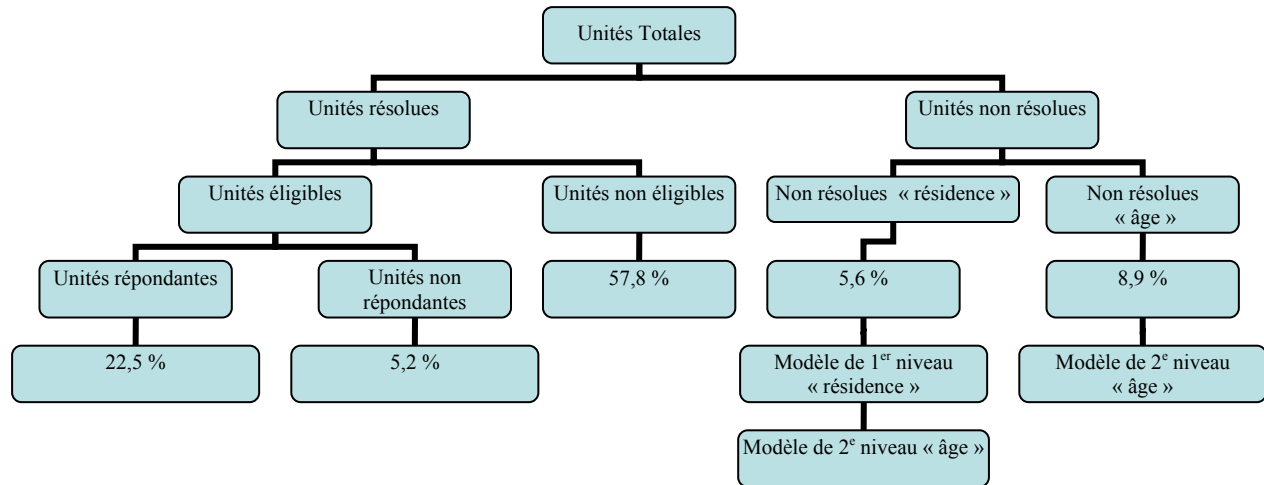
Pour tous les cas :	
1.	Est-ce que l'unité est éligible répondante?
→	Oui : Éligibilité « résidence » = Oui
→	Non : Étape 2
2.	Est-ce que l'unité a reçu un statut final sur le terrain ?
→	Oui :
→	Si le code de résultat final appartient à la catégorie « résidence » : Éligibilité « résidence » = Oui
→	Si le code de résultat final appartient à la catégorie « non résidence » : Éligibilité « résidence » = Non
→	Si le code de résultat final appartient à la catégorie « incapable de statuer » : Étape 4
→	Non : Étape 3
3.	Est-ce que l'unité est un numéro de téléphone au statut inconnu qui a atteint la limite d'appels de cinq et aucun contact n'a été détecté parmi ces cinq appels :
→	Oui : Éligibilité « résidence » = Non
→	Non : Étape 4
4.	Analyse du vecteur de paratonnerres : Existe-t-il dans la séquence de tentatives d'appels, un code de résultat qui appartient à la catégorie « résidence » ?
→	Oui : Éligibilité « résidence » = Oui
→	Non : Étape 5
5.	Analyse du vecteur de paratonnerres : Existe-t-il dans la séquence de tentatives d'appels, un code de résultat qui appartient la catégorie «non- résidence » ?
→	Oui : Éligibilité « résidence » = Non
→	Non : Étape 6
6.	Dans les données d'enquête, a-t-on pu confirmer un contact avec une résidence privée?
→	Oui : Éligibilité « résidence » = Oui
→	Non : Étape 7
7.	Dans les données d'enquête, a-t-on pu confirmer un contact avec une entité qui n'est pas une résidence?
→	Oui : Éligibilité « résidence » = Non
→	Non : Étape 8
8.	Est-ce qu'un code de résultat indiquant un télécopieur est obtenu pour plus de la moitié des codes de résultat de la séquence des appels?
→	Oui : Éligibilité « résidence » = Non
→	Non : Éligibilité « résidence » = Inconnue

Tableau 5.2-2**Résumé des règles déterministes utilisées pour la composante CA de l'ESG-21 – Éligibilité « âge »**

Pour tous les cas éligibilité « résidence »=Oui	
1.	Dans les données d'enquête : est-ce que la matrice de relation est complétée?
→	Oui : Y a-t-il au moins une personne de 45 ans et plus dans le ménage?
→	Oui : Éligibilité « âge » = Oui
→	Non : Éligibilité « âge » = Non
→	Non : Éligibilité « âge » = Inconnue

Suite à l'application des règles déterministes, il demeure deux catégories de cas non résolus. D'abord, les numéros de téléphone pour lesquels on ignore s'ils ont joint un ménage résidentiel (éligibilité « résidence » = inconnue selon le Tableau 5.2-1), et ensuite, ceux pour lesquels le numéro de téléphone a joint une résidence, mais qu'il est impossible de déterminer si le ménage est composé d'au moins un membre éligible (éligibilité « âge » = inconnu selon le Tableau 5.2-2). La Figure 5.2-1 illustre la répartition de ces cas et les étapes de modélisations. Parmi l'ensemble des unités totales, 5,6 % appartiennent à la première catégorie et 8,9 % à la seconde. Tel que mentionné à la section 5.1, la modélisation se fait en deux étapes : d'abord, l'assignation de l'éligibilité « résidence », ensuite, la détermination de l'éligibilité « âge ».

Figure 5.2-1
Répartition des unités au cycle 21 de l'ESG et étapes de modélisation



5.3 Modélisation de 1^{er} niveau - éligibilité « résidence »

Pour le modèle de 1^{er} niveau, l'assignation d'un statut d'éligibilité « résidence » se fait de façon aléatoire et indépendante pour chaque numéro de téléphone non résolu à l'intérieur d'un groupe homogène (GH), avec une certaine probabilité. Cette probabilité est donnée par π_{GH} où π_{GH} est le taux de succès « résidence » (c.-à-d. le nombre de numéros de téléphone ayant rejoint une résidence / nombre d'unités résolues) observé dans la partie résolue d'un groupe homogène de l'ESG-21.

Pour construire les GH, les sources d'information utilisées sont l'information tirée du plan de sondage et les parodonnées. À cette étape, l'information utilisée des parodonnées est le nombre total de tentatives sur le numéro de téléphone et l'information sur la vague de collecte. Nous avons retenu comme GH le résultat du croisement du statut du numéro de téléphone (répertorié, inconnu) avec la strate (urbaine, rurale), le nombre de tentatives (moins de 10 appels, 10 appels ou plus) et la vague de collecte. De plus, chaque groupe doit contenir au minimum de 20 unités et afficher une stabilité du paramètre π_{GH} par vague de collecte. Suite à l'application du modèle, 88 % des unités non résolues se sont vues assignées un statut de « résidence ». Elles se joignent donc aux autres unités non résolues « âge » et reçoivent un statut d'éligibilité « âge » via le modèle de 2^e niveau présenté dans la section suivante.

5.4 Modélisation de 2^e niveau - éligibilité « âge »

La première étape a fourni un ensemble de numéros de téléphone résidentiels. La seconde étape consiste en la détermination du statut d'éligibilité au niveau de l'âge de ces numéros de téléphone résidentiels, c.-à-d. « Quelle est la probabilité qu'une ligne téléphonique résidentielle rejoigne un ménage ayant au moins un membre de 45 ans et plus? ». La même approche que précédemment est utilisée: un statut sera assigné de façon aléatoire et indépendante pour chaque unité non résolue « âge » à l'intérieur d'un GH avec une certaine probabilité θ_{GH} . Pour le modèle d'éligibilité au niveau de l'âge, il a été décidé d'utiliser les strates comme GH, donc $GH=STR$. Quatre méthodes ont été examinées pour dériver θ_{STR} . Une fois ce paramètre obtenu et le modèle utilisé, le dénominateur du taux de réponse présenté à la formule 1 peut être calculé. La composante CA a permis d'obtenir 13 001 répondants.

Méthode 1

θ_{STR} est le taux de succès « âge » (c.-à-d. le nombre de numéros de téléphone ayant rejoint un ménage éligible / nombre de numéros de téléphone ayant rejoint une résidence) observé dans la partie résolue de l'ESG-21 par strate. Suite à l'application de ce modèle, un taux de réponse de 61,1 % est obtenu.

Autres méthodes

Nous avons aussi utilisé une autre variante où la proportion de lignes téléphoniques résidentielles qui joint un ménage appartenant à la population cible de la partie résolue et non résolue à l'intérieur de chaque strate correspond

à une autre source. Ceci nous permet de dériver un autre θ_{STR} qui sera appliqué à la partie non résolue. Selon cette logique, θ_{STR} est donné par:

$$\theta_{STR} = \max\left(\frac{n_{(résolus + non-résolus),STR} \theta_{\text{autre source},STR} - n_{résolus,STR} \theta_{résolus,STR}}{n_{non-résolus,STR}}, 0\right)$$

Nous avons utilisé les sources suivantes:

Méthode 2: La source externe est donnée par la proportion observée au cycle 20 de l'ESG. Ceci nous permet de travailler avec un bassin plus élevé de numéros de téléphone, où une proportion légèrement inférieure (7 %) de cas non résolus est observée. L'hypothèse est alors que la proportion observée est peut-être plus près de la réalité. Cette proportion est celle observée parmi les numéros de téléphone, donc non pondérée. Le taux de réponse obtenu avec cette méthode est 61,1 %, exactement comme avec la méthode 1.

Méthode 3: La source externe est donnée par la proportion représentant la moyenne historique des trois derniers cycles de l'ESG. L'idée est d'avoir un bassin encore plus élevé de numéros de téléphone, mais moins d'actualité, puisqu'il contient des données retournant jusqu'en 2004. Cette proportion est celle observée suite au processus de collecte, toujours non pondérée. Cette méthode donne un taux de réponse de 63,4 %.

Méthode 4: La source externe est donnée par la proportion provenant de l'Enquête sur le service téléphonique résidentielle (ESTR) de 2007. Pour chaque strate de l'ESG, nous avons produit l'estimation de la proportion du nombre de ménages ayant au moins un membre de 45 ans et plus pouvant être joint au moyen d'une ligne téléphonique résidentielle parmi tous les ménages pouvant être joint par une ligne téléphonique résidentielle. Cette méthode permet d'obtenir un taux de réponse de 59,5 %.

L'option retenue est la méthode 2. Elle donne le même résultat au niveau national que l'option 1 et des résultats très similaires par strate. Il en va de même avec l'option 4. En calculant des taux de réponse minimum et maximum, basés respectivement sur l'estimation selon la borne supérieure et inférieure de l'estimation provenant de l'ESTR, on obtient un taux de réponse minimum de 57,3 % et maximum de 62 %.

6. Taux de réponse de la composante suivi et taux de réponse global

Tel que décrit à la section 2, l'autre composante de l'ESG-21 provient d'un suivi de l'ESG-20. Dans le cas d'une enquête de suivi, les renseignements sur la non-réponse doivent être fondés sur la non-réponse cumulative de l'enquête principale et de l'enquête secondaire. La méthode retenue pour cette composante de l'ESG-21 est simplement le produit des deux taux de réponse: celui de l'ESG-20 à l'estimation et celui obtenue à la collecte de l'ESG-21. On retrouve donc : 67,4 % x 80,1 % = 54 %. Il est à noter qu'une des hypothèses sous-jacente à ce calcul est que le taux de réponse au cycle 20 chez les 45 ans et plus est le même que celui chez les gens de moins de 45 ans.

Le taux de réponse global pour les deux composantes est obtenu par :

$$\text{Taux de réponse} = \frac{\text{Unités répondantes}_{CA} + \text{Unités répondantes}_{\text{suivi}}}{(\text{Unités éligibles} + \text{Unités estimées éligibles})_{CA} + (\text{Unités éligibles} + \text{Unités estimées éligibles})_{\text{suivi}}}$$

Il est important de noter que dans cette formule, le nombre d'unités éligibles et d'unités estimées éligibles provenant de la composante suivi du cycle 20, doit être estimé. En effet, l'ESG-20 ciblant les personnes âgées de 15 ans et plus, il est possible que certains ménages n'aient pas au moins une personne de 45 ans et plus. Il fut décidé de simplement diviser le nombre de répondants de l'ESG-21 de la composante de suivi (10 403) par le taux de réponse de cette composante obtenu à l'étape précédente (54 %). Le taux de réponse global ainsi calculé est de 57,7 %. Une autre approche aurait été de modéliser les cas non résolus du cycle 20 en utilisant une approche similaire à celle présentée à la section 5.

7. Conclusion

Le présent article permet d'illustrer une autre utilisation des parodonnées. En plus d'être des données essentielles pour la gestion de la collecte et aider à informer et à comprendre la non-réponse, elles sont aussi des données clés pour le calcul du taux de réponse, en particulier pour une enquête CA. Le taux de réponse est une mesure importante de la qualité étant la première étape à franchir pour donner un ordre de grandeur à la non-réponse d'une enquête. C'est un indicateur complexe à dériver dans une enquête CA et notamment dans une enquête CA ciblant un sous-ensemble d'une population. Les lignes directrices de Statistique Canada donnent un cadre de travail qui doit être adapté en fonction des caractéristiques des enquêtes. L'ESG-21 est un bon exemple démontrant les nombreuses alternatives possibles à chaque étape du calcul du taux de non-réponse. Tout en restant à l'intérieur des lignes directrices, le choix des différentes règles déterministes, groupes homogènes et paramètres dans les modèles, approches ou méthodes, affecte la comparabilité entre les enquêtes qui auraient choisi des avenues différentes. Dans certains cas, et en particulier pour l'ESG-21, le taux de réponse est un paramètre estimé. Ceci illustre les limites à utiliser qu'un seul indicateur pour résumer la qualité d'une enquête complexe.

Bibliographie

- Kish, L. (1976). Optima and Proxima in Linear Sample Designs, *Journal of the Royal Statistical Society A*, 139, 80-95.
- Laflamme, F., Maydan, M. et Miller, A. (2008). Using Paradata to Actively Manage Data Collection Survey Process, *Proceedings of the Section on Survey Research Methods*, American Statistical Association.
- Statistique Canada (2000). Politique visant à informer les utilisateurs sur la qualité des données et la méthodologie. Manuel des politiques, 2.3, Statistique Canada
- Statistique Canada (2001). Normes et lignes directrices pour la déclaration des taux de non-réponse. Rapport technique de Statistique Canada.