

L'utilité d'utiliser les enquêtes en ligne pour fournir des estimations officielles concernant les principaux résultats en matière de santé, une étude pilote

Yulei He, Hee-Choon Shin, Bill Cai, Vladislav Beresovsky, Paul Scanlon, Van Parsons, Peter Meyer et Jennifer Parker¹

Résumé

Étant donné la hausse vertigineuse des coûts et des contraintes en matière de ressources pour les enquêtes traditionnelles, les enquêtes en ligne sont de plus en plus utilisées comme solution de rechange moins chère. Toutefois, la documentation publiée antérieurement a bien documenté les limites générales de l'utilisation des enquêtes en ligne. De même, il existe également des recherches sur la compréhension et l'amélioration de leur utilité. Du point de vue d'un organisme statistique de santé, une importante question demeure : les enquêtes en ligne peuvent-elles être utilisées pour produire des estimations nationales officielles des principaux résultats en matière de santé (p. ex., prévalence du diabète, couverture d'assurance)? C'est pourquoi le National Center for Health Statistics (NCHS) mène une série d'études sur les enquêtes en ligne. Le double but de ces études est d'évaluer l'utilité des enquêtes en ligne dans la production d'estimations officielles des soins de santé et d'évaluer leur utilisation afin d'examiner les structures de réponses aux questions. Ces études comprennent également les données provenant de la National Health Interview Survey (NHIS), qui est une importante source de données utilisées pour produire des estimations officielles en matière de santé pour l'ensemble de la population des États-Unis. Ce document présente le contexte de ce projet et certaines comparaisons initiales entre les estimations de l'enquête en ligne et celles de la NHIS. Les résultats ont révélé que les estimations pondérées des résultats en matière de santé de l'échantillon en ligne ont tendance à être pires que celles de la NHIS. Les différences apparentes des estimations entre l'enquête en ligne et la NHIS soulèvent des préoccupations en ce qui concerne l'utilisation d'enquêtes en ligne pour remplacer les enquêtes traditionnelles afin de produire des estimations officielles des résultats en matière de santé en général.

Mots-clés : Enquête sur la santé; National Health Interview Survey; statistiques officielles; panel probabiliste; enquête en ligne.

1. Introduction

Les enquêtes par échantillonnage sont une approche importante pour recueillir de l'information auprès de sujets humains à des fins de recherche scientifique. Dans le passé, l'approche par enquête a principalement reposé sur trois méthodes de collecte de données : les interviews en personne, les interviews téléphoniques et les enquêtes envoyées par la poste. Pendant de nombreuses années, ces méthodes traditionnelles d'enquête ont été efficaces pour recueillir des renseignements sur la santé et ont fourni des estimations nationales officielles sur l'état de santé. Des exemples de réussite comprennent la National Health Interview Survey (NHIS), qui est fondée sur des interviews en personne menées par le National Center for Health Statistics (NCHS). La NHIS est la principale source de données en vue de fournir des statistiques officielles sur d'importants résultats en matière de santé, comme la couverture d'assurance-maladie et la prévalence des maladies graves (p. ex. diabète). On peut trouver d'autres renseignements à <http://www.cdc.gov/nchs/nhis.htm> (en anglais) par exemple.

Toutefois, toutes les grandes enquêtes qui font appel aux méthodes traditionnelles de collecte de données sont touchées par la baisse des taux de réponse (Williams et Brick, 2017). Une des principales raisons de cette diminution pourrait être la difficulté croissante de communiquer avec les unités d'échantillonnage et de les convaincre de participer aux enquêtes (Singer, 2006). L'augmentation des taux de non-réponse soulèverait naturellement des préoccupations de biais attribuables à la non-réponse pour les estimations de l'enquête (Tourangeau, 2017). De plus, pour les enquêtes en personne sur la santé, à grande échelle, l'augmentation du coût unitaire apporte d'autres défis. Étant donné le manque croissant de ressources et la réduction des taux de réponse, les chercheurs et administrateurs d'enquêtes sont obligés de trouver une autre approche économique pour recueillir des renseignements sur l'état de santé de la population tout en maintenant la rigueur scientifique et méthodologique des enquêtes probabilistes.

¹Yulei He, wdq7@cdc.gov Division of Research and Methodology, National Center for Health Statistics, US Centers for Disease Control and Prevention, 3311 Toledo Road, Hyattsville, MD, 20782, U.S.A.

Au cours des 20 dernières années environ, le domaine de la méthodologie de l'enquête a connu une expansion innovante et stimulante — les enquêtes en ligne (p. ex. Tourangeau et coll. (2013), Baker et coll. (2013) et les références qu'ils contiennent). Le développement d'Internet et du Web a fondamentalement changé la structure de nos voies de communication quotidiennes. L'échange de courrier électronique (courriels) et l'envoi de messages instantanés sur le Web sont maintenant considérés comme des activités ordinaires dans la plupart des pays industrialisés, y compris les États-Unis. Il semble être un choix naturel de tirer profit de la technologie fondée sur le courriel et Internet afin d'élaborer, d'administrer et de mettre en œuvre des enquêtes en ligne. Non seulement les enquêtes en ligne représentent-elles un progrès dans l'évolution des questionnaires à remplir soi-même, mais leur coût global est bien moindre que les méthodes traditionnelles de collecte de données, comme les interviews en personne, pour lesquelles chaque cas rempli nécessite un contact avec un intervieweur, ce qui suppose des coûts élevés liés à la rémunération des intervieweurs et à la logistique.

Dans le secteur privé, les enquêtes en ligne ont été facilement appliquées dans les sondages de marketing économique ou d'opinion. Les enquêtes en ligne ont souvent été utilisées dans les enquêtes menées auprès d'entreprises ou d'établissements et ont remplacé ou complété les enquêtes postales. Leur utilisation augmente également dans les enquêtes menées par le gouvernement. Elles sont de plus en plus utilisées pour les recensements de la population (p. ex. les enquêtes en ligne sont utilisées comme mode de collecte de données pour l'American Community Survey menée par le Census Bureau des États-Unis). Les organismes statistiques ont étudié l'utilisation d'enquêtes en ligne dans le cadre d'une stratégie à mode mixte. Par exemple, en 2013, le NCHS a effectué une enquête de suivi par téléphone/en ligne pour un sous-ensemble de répondants adultes de l'échantillon de la NHIS de 2012.

Toutefois, malgré l'attrait au niveau budgétaire et l'aspect potentiellement pratique, la recherche actuelle a documenté des inconvénients inhérents aux enquêtes en ligne, qui peuvent mener à diverses formes d'erreurs. Une des grandes préoccupations, c'est que le faible coût des enquêtes en ligne reflète souvent une indifférence presque complète à l'égard des principes d'échantillonnage probabiliste. Pourtant, ces principes forment la base de la plupart des enquêtes de haute qualité (p. ex. NHIS) qui sont conçues pour faire des inférences à de vastes populations et produire des statistiques officielles par l'entremise d'organismes statistiques. En d'autres mots, de nombreuses enquêtes en ligne sont fondées sur des échantillons de commodité, ce qui fait qu'il est souvent difficile de déterminer si l'échantillon tiré des enquêtes en ligne est représentatif de la population cible, ce qui remet d'autant plus en question l'exactitude des estimations des enquêtes en ligne. Cette observation est fondée sur bon nombre de documents comparant les estimations provenant d'enquêtes en ligne avec celles provenant d'enquêtes probabilistes traditionnelles (p. ex. Lee et Valiant, 2009; Bethlehem, 2010).

Plus précisément, en pratique, la plupart des enquêtes en ligne étaient fondées sur des panels volontaires (autosélectionnés/volontaires) qui n'ont pas la représentativité des échantillons probabilistes conventionnels (Couper, 2000). Malgré la préoccupation générale au sujet de la représentativité des enquêtes en ligne de volontaires, de vastes recherches ont été menées sur le rendement de l'utilisation de ces enquêtes en ligne pour mesurer les mesures de la santé. Couper et coll. (2007) ont recensé plusieurs projets de recherche utilisant les enquêtes en ligne pour recueillir des renseignements sur divers sujets liés à la santé dans l'ensemble de multiples domaines cliniques. On peut trouver des documents plus récents dans Erens et coll. (2014) et Nelson et coll. (2014), par exemple. Ces études ont également été ciblées à des mesures de santé précises et ont souvent été menées à petite échelle. La nature volontaire de ces enquêtes en ligne ainsi que leurs portées limitées ont imposé certaines limites aux conclusions de l'étude et les ont rendues moins généralisables.

À titre de principal organisme statistique de santé des États-Unis, le NCHS s'intéresse à la faisabilité de l'utilisation d'enquêtes en ligne pour la prestation de statistiques officielles nationales sur un large éventail de résultats en matière de santé. Il s'agit là d'un très vaste problème de recherche et c'est pourquoi nous avons mené une série d'études sur les enquêtes en ligne à des fins d'investigation. L'objectif spécifique de ce document est de décrire le contexte et la mise en œuvre de la première étude sur les enquêtes en ligne et de fournir des statistiques descriptives. Les estimations de l'enquête en ligne sont comparées à celles de la NHIS afin de fournir un aperçu de la possibilité de les utiliser pour des estimations officielles. Le reste du document est organisé comme suit : à la section 2, on décrit la conception et la mise en œuvre de l'enquête en ligne; à la section 3, on présente les estimations et l'étude comparative avec la NHIS; à la section 4, on résume les constatations et expose les futurs plans de recherche.

2. Plan de l'étude

Dans le présent document, l'enquête en ligne est désignée sous le nom de Research and Development Survey (RANDS) (enquête sur la recherche et le développement). Notre étude comprend la description et la comparaison des estimations sur la santé de la RANDS aux estimations sur la santé correspondantes de la NHIS, qui offre un aperçu fiable de l'état de santé de la population générale des États-Unis.

2.1 Panel Gallup en ligne

En général, les données d'enquêtes en ligne peuvent être recueillies par échantillonnage probabiliste ou échantillonnage non probabiliste. Couper (2000) a classé les enquêtes en ligne en plusieurs types selon le mode de collecte des échantillons. Toutefois, selon le consensus retrouvé dans la littérature existante, les estimations tirées d'enquêtes non probabilistes fondées sur des panels volontaires n'auraient peut-être pas des propriétés statistiques souhaitables (Baker et coll., 2013). C'est-à-dire que leur représentativité de la population serait douteuse. Cela a mené à inclure uniquement les enquêtes en ligne fondées sur un panel probabiliste pour ce projet.

Les données de l'échantillon ont été recueillies à l'aide des enquêtes en ligne par panel Gallup. Le panel Gallup est un panel de recherche qui est représentatif de l'ensemble de la population américaine. Plus précisément, au moment de la collecte des données de la RANDS, le panel Gallup a choisi certains membres potentiels en utilisant la composition aléatoire de téléphones conventionnels et de téléphones cellulaires, ou un échantillonnage fondé sur les adresses pour communiquer avec les ménages américains de façon aléatoire. Après avoir recruté une personne pour se joindre au panel, en fonction de son statut d'accès à Internet, Gallup a communiqué avec elle pour remplir des enquêtes par courriel ou par la poste. Gallup a également communiqué avec les membres par téléphone. La capacité de tirer parti de plusieurs modes d'enquêtes afin de cibler une population particulière n'est pas réalisable dans le cadre de panels volontaires. Il est possible de trouver les détails du panel Gallup utilisé pour la RANDS dans Rookey et coll. (2008).

2.2 Plan d'échantillonnage et collecte de données

La première vague de la RANDS a été conçue au départ pour compter un échantillon de 2 000 personnes avec un taux de réponse de 40 %. L'échantillonnage aléatoire a été utilisé dans plusieurs strates définies par la race et l'origine ethnique (blanc seulement non hispanique, noir seulement non hispanique, autre non hispanique, hispanique), l'âge (18 à 34 ans, 35 à 54 ans, 55 ans et plus), et le niveau de scolarité (diplôme d'études secondaires ou moins, études collégiales partielles, diplômes collégiaux). La première vague de collecte de données a commencé le 11 février 2015 et s'est terminée le 12 septembre 2015. En raison du taux de réponse plus faible que prévu, de nombreux rappels ont été envoyés aux participants à l'enquête. Au total, 9 809 répondants ont été invités à participer à l'enquête. Lorsque l'enquête a été retirée du champ le 12 septembre 2015, un total de 2 304 répondants avaient rempli le questionnaire, pour un taux de réponse global de 23,5 %. Il est possible de trouver des renseignements plus précis dans (les références sur le travail sur le terrain de Gallup).

2.3 Conception du questionnaire

Afin de permettre la comparaison entre la NHIS et la RANDS, certaines questions de la NHIS ont été utilisées dans le questionnaire de la RANDS. Après avoir consulté des spécialistes du domaine, un sous-ensemble de 88 questions de la NHIS (provenant principalement du questionnaire-échantillon destiné aux adultes de la NHIS) a été utilisé. Des renseignements détaillés sur les questions utilisées dans la RANDS sont disponibles sur demande.

3. Analyse des données

Pour la RANDS, Gallup Inc. a fourni des poids de sondage au NCHS qui ont tenu compte de l'ensemble de la composition du panel et de la réponse à l'enquête par le biais de la poststratification et d'un ajustement de non-réponse. Les poids de sondage de la RANDS ont été normalisés de sorte que la taille de l'échantillon pondéré soit égale à la taille de l'échantillon non pondéré de répondants (n=2 304). Tout au long du document, nous traitons la RANDS comme une enquête probabiliste et utilisons ses poids de sondage pour l'estimation.

3.1 Utilisation des données de 2015 de la NHIS à des fins de comparaison

Les principales données comparatives (c.-à-d. l'enquête de référence) ont été établies à partir de la NHIS. Plus précisément, un sous-ensemble des données d'enquête de la NHIS recueillies d'octobre à décembre 2015 a été extrait (n=7 723). Les poids de sondage correspondant de la NHIS sont ciblés pour fournir des estimations annuelles.

3.2 Données démographiques

Tout au long du document, la comparaison statistique est fondée sur des données d'enquête ajustées sur des tests de chi-carré (c.-à-d., tests de Rao-Scott, Scott et Rao, 1981). Toutes les estimations sont fondées sur les données observées, car les taux de non-réponse partielle sont assez faibles (p. ex. moins de 5 %) en général. La variable de revenu a une non-réponse partielle élevée pour les deux enquêtes.

Le tableau 3.2-1 montre la répartition pondérée de plusieurs variables démographiques dans la RANDS et la NHIS. Il y avait des différences selon le niveau de scolarité, le revenu, l'état matrimonial et la race-ethnicité. Les estimations de la RANDS montrent des pourcentages plus élevés pour sans diplôme d'études secondaires ou diplôme d'études secondaires. La RANDS comporte également plus de non-réponses à la question sur le revenu. Tout particulièrement, par rapport aux estimations de la NHIS, les estimations de la RANDS indiquent un pourcentage plus élevé d'adultes blancs non hispaniques et un plus faible pourcentage d'Asiatiques non hispaniques et d'autres groupes.

Tableau 3.2-1
Répartition des groupes démographiques dans la RANDS et la NHIS, vague I

Variable	RANDS		NHIS	
	Moyenne	ET	Moyenne	ET
Âge				
18 à 36 ans	29,5 %	0,91 %	29,6 %	0,87 %
35 à 55 ans	33,7 %	0,99 %	34,4 %	0,82 %
55 à 65 ans	17,0 %	0,92 %	16,7 %	0,59 %
65 à 75 ans	13,4 %	0,85 %	11,3 %	0,49 %
> 75 ans	6,4 %	0,62 %	8,0 %	0,41 %
Sexe				
Femmes	51,1 %	1,49 %	51,8 %	0,79 %
Hommes	48,9 %	1,49 %	48,2 %	0,79 %
Région de recensement				
Nord-Est	18,1 %	1,18 %	18,6 %	0,84 %
Midwest	21,3 %	1,20 %	22,4 %	0,89 %
Sud	36,8 %	1,44 %	36,8 %	1,01 %
Ouest	23,8 %	1,22 %	22,2 %	0,92 %

Niveau de scolarité*				
Sans diplôme d'études secondaires	3,6 %	0,69 %	15,3 %	0,66 %
Diplômes d'études secondaires	37,0 %	1,05 %	20,9 %	0,73 %
Quelques études collégiales, grade d'associé (formation professionnelle)	29,2 %	0,92 %	30,4 %	0,76 %
Baccalauréat ou niveau supérieur	30,3 %	0,90 %	33,4 %	0,96 %
Revenu*				
Données manquantes	22,9 %	1,28 %	8,6 %	0,48 %
< 50 000 \$	29,5 %	1,37 %	41,0 %	0,91 %
50 000 \$ à 99 999 \$	26,4 %	1,26 %	26,5 %	0,75 %
> 100 000 \$	21,2 %	1,16 %	23,9 %	0,88 %
État matrimonial*				
Marié/vit avec un conjoint	62,4 %	1,46 %	60,4 %	0,84 %
Célibataire	23,2 %	1,25 %	22,1 %	0,74 %
Séparé/Divorcé/Veuf	14,4 %	1,03 %	17,5 %	0,57 %
Groupes raciaux et ethniques*				
Blancs non hispaniques	72,4 %	0,93 %	65,3 %	1,00 %
Noirs non hispaniques	11,5 %	0,71 %	12,1 %	0,62 %
Asiatiques non hispaniques	1,0 %	0,20 %	5,9 %	0,35 %
Hispaniques	14,8 %	0,74 %	15,7 %	0,76 %
Autres	0,3 %	0,08 %	1,1 %	0,18 %

Note : ET signifie « erreur-type »; * indique un niveau significatif de 5 % d'après le test du chi-carré comparant la RANDS et la NHIS.

3.3 Principaux indicateurs sur la santé

Le tableau 3.3-1 présente certaines statistiques descriptives de variables choisies qui sont couramment utilisées pour résumer l'accès aux soins de santé, l'utilisation de services de santé, l'état de santé et des comportements connexes liés à la santé. Il existe des différences significatives entre les estimations de la RANDS et celles de la NHIS pour certains des principaux indicateurs de la santé. Même si les estimations des participants aux enquêtes en ligne de la RANDS ont indiqué qu'un pourcentage élevé d'adultes sont couverts par l'assurance-maladie (plus de 90 %) et qu'ils utilisent beaucoup l'ordinateur pour obtenir des renseignements sur la santé (plus de 80 %), les estimations de la RANDS en ce qui concerne la difficulté à obtenir des services de soins de santé abordables et en temps opportun étaient plus élevées par rapport à celles de la NHIS. La RANDS a également obtenu des estimations plus élevées que la NHIS pour des éléments comme le fait d'être obèse, de présenter davantage de prédiabète, d'asthme et de maladies pulmonaires, d'éprouver plus de problèmes nerveux et émotionnels, et d'être plus préoccupés par certains des moyens de subsistance de première nécessité comme la nourriture.

Tableau 3.3-1

Répartition des principaux indicateurs sur la santé dans la RANDS et la NHIS, vague I

Variable	RANDS		NHIS	
	Moyenne	ET	Moyenne	ET

Êtes-vous couvert par une forme quelconque d'assurance-maladie ou tout autre type de régime de soins médicaux?*	93,4 %	0,80 %	90,5 %	0,51 %
Au cours des 12 derniers mois, vous est-il arrivé de ne pas pouvoir vous permettre et de ne pas recevoir n'importe lequel des services de soins de santé?*	31,5 %	1,41 %	15,9 %	0,57 %
Avez-vous reporté l'obtention de soins pour une raison ou une autre au cours des 12 derniers mois?	26,1 %	1,32 %	10,8 %	0,44 %
Obésité*				
Insuffisance pondérale (IMC < 18,5)	1,2 %	0,35 %	1,7 %	0,19 %
Poids normal (IMC >=18,5)	30,9 %	1,40 %	34,9 %	0,74 %
Surcharge pondérale (IMC >=25)	33,8 %	1,43 %	33,9 %	0,67 %
Obésité (IMC >=30)	34,1 %	1,41 %	29,5 %	0,77 %
Mis à part pendant la grossesse, avez-vous déjà été informé(e) par un médecin que vous souffrez de diabète ou de diabète sucré?*				
Non	81,2 %	1,12 %	85,6 %	0,51 %
Prédiabète	9,8 %	0,89 %	5,3 %	0,35 %
Oui	9,0 %	0,79 %	9,1 %	0,42 %
Avez-vous déjà été informé(e) par un médecin que vous souffrez d'hypertension (tension artérielle élevée)?	32,0 %	1,26 %	30,6 %	0,75 %
Avez-vous déjà été informé(e) par un médecin que vous souffrez d'asthme?*	17,2 %	1,14 %	12,9 %	0,53 %
Souffrez-vous d'emphysème, de bronchopneumopathie chronique obstructive ou de bronchite chronique?*	9,5 %	0,89 %	6,2 %	0,39 %
À quelle fréquence fumez-vous actuellement des cigarettes? Tous les jours, certains jours, ou pas du tout?				
Pas du tout.	87,3 %	0,99 %	85,7 %	0,55 %
Parfois	4,1 %	0,59 %	3,6 %	0,26 %
Chaque jour	8,6 %	0,84 %	10,7 %	0,52 %
Au cours d'une année donnée, avez-vous bu au moins 12 verres d'une boisson alcoolisée de quelque type que ce soit?*	70,3 %	1,37 %	64,4 %	0,81 %
Au cours des 30 derniers jours, vous sentiez-vous tout le temps ou la plupart du temps triste/nerveux(se)/agité(e)/désespéré(e)/aucun effort/sans valeur?*	18,4 %	1,19 %	13,5 %	0,60 %
À quelle fréquence (nombre de fois par semaine) faites-vous une combinaison d'activités de loisirs? (Échelle de 0 à 7)*	3,8190	0,0732	2,9496	0,0497
État de santé autodéclaré*				
Excellent état de santé	12,7 %	0,99 %	28,1 %	0,70 %
Très bon	41,5 %	1,48 %	31,6 %	0,76 %
Bon	34,2 %	1,40 %	27,2 %	0,70 %
Passable	9,6 %	0,87 %	10,2 %	0,49 %
Mauvaise	2,0 %	0,41 %	2,9 %	0,23 %

Je craignais de manquer de nourriture avant d'avoir de l'argent pour en acheter d'autres*.				
Souvent	5,0 %	0,71 %	4,0 %	0,31 %
Parfois	15,8 %	1,12 %	9,8 %	0,46 %
Jamais	79,3 %	1,24 %	86,1 %	0,56 %
Avez-vous un emploi rémunéré ou une entreprise?	60,4 %	1,37 %	60,2 %	0,83 %
Au cours des 12 derniers mois, avez-vous utilisé un ordinateur pour rechercher des renseignements sur la santé dans Internet?*	82,8 %	1,14 %	52,9 %	0,90 %

Note : Et signifie « erreur-type »; * indique un niveau significatif de 5 % d'après le test du chi-carré comparant la RANDS et la NHIS.

3.4 Autres variables

En raison de la brièveté de la présentation, nous n'examinons pas les estimations de toutes les questions de l'enquête. Elles sont disponibles sur demande. Les principaux indicateurs de la santé présentés dans le tableau 3.3.1 ont agi à titre de porte d'entrée ou de résumé des questions dans le cadre de l'enquête. Les estimations des questions de suivi correspondantes en général ont montré des modèles de comparaison similaires, fournis plus en détail. Dans l'ensemble, par rapport aux estimations de la NHIS, les estimations de la RANDS indiqueraient que les adultes sont en moins bonne santé, ont plus de problèmes affectifs, ont des comportements plus risqués, comme la consommation d'alcool, et ont plus de difficulté à avoir accès à des services de soins de santé abordables et en temps opportun.

4. Conclusion

D'après nos connaissances, la RANDS est peut-être la première utilisation d'une enquête probabiliste par panel en ligne pour une liste exhaustive d'indicateurs de la santé, et menée par un organisme statistique, malgré des études approfondies sur l'utilisation d'enquêtes en ligne par panels de volontaires. Le présent document décrit certains renseignements généraux et fournit des analyses descriptives des données de la vague 1 de la RANDS. Notre étude montre que les enquêtes en ligne peuvent produire des estimations comparables à celles de l'enquête traditionnelle en personne (c.-à-d. NHIS) pour certains des éléments démographiques et liés à la santé (p. ex. hypertension). Cependant, pour de nombreux autres éléments clés de la santé, les estimations directes de l'enquête en ligne sont considérablement différentes de celles de l'enquête traditionnelle. Cette constatation soulève des préoccupations au sujet de la faisabilité de l'utilisation d'estimations directes provenant d'enquêtes probabilistes par panel en ligne à titre de statistiques officielles des résultats généraux sur l'état de santé.

Les principales raisons de la différence apparente pourraient inclure l'erreur d'échantillonnage et l'erreur de couverture. Même si nous avons délibérément choisi d'utiliser le panel Gallup, qui est un panel probabiliste généré par la composition aléatoire, ses propriétés de couverture sont moins claires par rapport aux enquêtes nationales bien établies comme la NHIS. Le taux élevé de non-réponse (plus de 70 % en général) de la RANDS peut également biaiser les estimations. Nous notons que Gallup Inc. a utilisé certains ajustements traditionnels de poststratification et de non-réponse par certains groupes démographiques. La prochaine étape de notre recherche consiste à étudier la faisabilité de l'utilisation de techniques de modélisation de pointe pour caler les estimations de l'enquête en ligne. Il est possible de trouver une partie de la documentation connexe pour le calage dans Lee (2006a, 2006b) et Lee et Valliant (2009). Même si elles ont été conçues pour des données non probabilistes, nous nous attendons à ce que ces approches et d'autres approches connexes pour le calage à une « référence » s'appliquent directement à une variété de données calées à la NHIS. Dans notre contexte, puisque notre intérêt met l'accent sur les résultats en matière de santé, les totaux de contrôle nécessaires dans les méthodes courantes de calage peuvent dépasser la démographie typique et être élargis aux variables en matière de santé, qui peuvent facilement être estimées à partir de la NHIS, l'enquête de référence.

Bibliographie

- Baker R. P., J. M. Brick, N. A. Bates, M. Battaglia, M. J. Couper, J. A. Dever, K. J. Gile, et R. Tourangeau (2013), « Summary Report of AAPOR Task Force on Non-probability Sampling », *Journal of Survey Statistics and Methodology*, 1, p. 90-143.
- Bethlehem J. (2010), « Selection Bias in Web Surveys », *International Statistical Review*, 78, p. 161-188.
- Cantrell M. A., et P. Lupinacci (2007), « Methodological Issues in Online Data Collection », *Journal of Advanced Nursing*, 60, p. 544-549.
- Couper M. P. (2000), « Web Surveys: A Review of Issues and Approaches », *Public Opinion Quarterly*, 64, p. 464-494.
- Couper M. P. (2007), « Issues of Representation in eHealth Research (with a Focus on Web Surveys) », *American Journal of Preventive Medicine*, 32, p. S83-S89.
- Dever J. A., A. Rafferty, et R. Valliant (2008), « Internet Surveys: Can Statistical Adjustments Eliminate Coverage Bias? », *Survey Research Methods*, 2, p. 47-62.
- Erens B., S. Burkill, M. P. Couper, F. Conard, S. Clifton, C. Tanton, A. Phelps, J. Datta, C. H. Mercer, P. Sonnenberg, P. Prah, K. R. Mitchell, K. Wellings, A. M. Johnson, et A. J. Copas (2014), « Nonprobability Web Surveys to Measure Sexual Behaviors and Attitudes in the General Population: A Comparison with a Probability Sample Interview Survey », *Journal of Medical Internet Research*, 16, p. e276.
- Gallup Inc. (2018), *CDC/NCHS/RANDS Field Work Summary*.
- Greby S. (2014), *Surveys to Monitor Influenza Coverage among Pregnant Women Final Report*, Abt Associates. Contract # GS-10F-0086K.
- Lee S. (2006a), « An Evaluation of Nonresponse and Coverage Errors in a Prerecruited Probability Web Panel Survey », *Social Science Computer Review*, 24, p. 460-475.
- Lee S. (2006b), « Propensity Score Adjustment as a Weighting Scheme for Volunteer Panel Web Surveys », *Journal of Official Statistics*, 22, p. 329-349.
- Lee S., et R. Valliant (2009), « Estimation for Volunteer Panel Web Surveys Using Propensity Score Adjustment and Calibration Adjustment », *Sociological Methods and Research*, 37, p. 319-343.
- McCabe S. E., C. J. Boyd, M. P. Couper, D. Crawford, et H. D'Arcy (2002), « Mode Effects for Collecting Alcohol and Other Drug Use Data: Web and US Mail », *Journal of Studies on Alcohol and Drugs*, 63, p. 755-761.
- Nelson E.J., J. Hughes, J. M. Oakes, J. S. Pankow, et S. L. Kulasingam (2014), « Estimation of Geographic Variation in Human Papillomavirus Vaccine Uptake in Men and Women: An Online Survey Using Facebook Recruitment », *Journal of Medical Internet Research*, 16, p. e198.
- Rookey B.D., S. Hanway, et D. A. Dillman (2008), « Does a Probability-Based Household Panel Benefit from Assignment to Postal response as an Alternative to Internet-Only? », *Public Opinion Quarterly*, 72, p. 962-984.
- Scanlon P. (2018), *Using Targeted Embedded Probes to Quantify Cognitive Interviewing Findings*.
- Schonlau M., A. van Soest, A. Kapteyn, et M. P. Couper (2009), « Selection Bias in Web Surveys and the Use of Propensity Score », *Sociological Methods and Research*, 37, p. 291-318.
- Scott A.J., et J. N. K. Rao (1981), *Chi-square Tests for Homogeneity with Proportions Estimated from Survey Data*, New York: Academic Press.

Singer, E. (2006), « Nonresponse Bias in Household Surveys », *Public Opinion Quarterly*, 70(5), p. 637–645.

Tourangeau R., G. G. Frederick, et M. P. Couper (2013), *The Science of Web Surveys*, Oxford: Oxford University Press.

Tourangeau R. (2017), « Paradoxes of Nonresponse », *Public Opinion Quarterly*, 81, p. 803-814.

Williams D., et M. J. Brick (2018), « Trends in U.S. Face-to-Face Household Survey Nonresponse and Level of Effort », *Journal of Survey Statistics and Methodology*, 6, p. 186-211.