

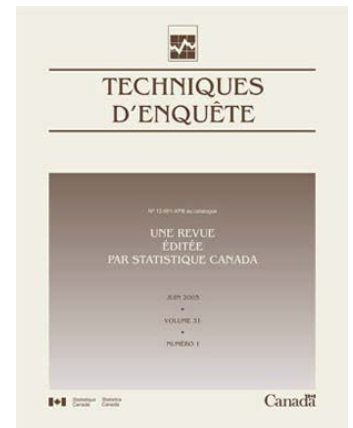
N° 12-001-X au catalogue  
ISSN 1712-5685

## Techniques d'enquête

# Commentaires à propos de l'article de Rao et Fuller (2017)

par Sharon L. Lohr

Date de diffusion : le 21 décembre 2017



Statistique  
Canada

Statistics  
Canada

Canada

---

## Comment obtenir d'autres renseignements

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

Vous pouvez également communiquer avec nous par :

**Courriel** à [STATCAN.infostats-infostats.STATCAN@canada.ca](mailto:STATCAN.infostats-infostats.STATCAN@canada.ca)

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

- Service de renseignements statistiques 1-800-263-1136
- Service national d'appareils de télécommunications pour les malentendants 1-800-363-7629
- Télécopieur 1-877-287-4369

### Programme des services de dépôt

- Service de renseignements 1-800-635-7943
- Télécopieur 1-800-565-7757

## Normes de service à la clientèle

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

## Note de reconnaissance

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

## Signes conventionnels dans les tableaux

Les signes conventionnels suivants sont employés dans les publications de Statistique Canada :

- . indisponible pour toute période de référence
- .. indisponible pour une période de référence précise
- ... n'ayant pas lieu de figurer
- 0 zéro absolu ou valeur arrondie à zéro
- 0<sup>s</sup> valeur arrondie à 0 (zéro) là où il y a une distinction importante entre le zéro absolu et la valeur arrondie
- <sup>p</sup> provisoire
- <sup>r</sup> révisé
- x confidentiel en vertu des dispositions de la *Loi sur la statistique*
- <sup>E</sup> à utiliser avec prudence
- F trop peu fiable pour être publié
- \* valeur significativement différente de l'estimation pour la catégorie de référence ( $p < 0,05$ )

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2017

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

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

*This publication is also available in English.*

---

## Commentaires à propos de l'article de Rao et Fuller (2017)

Sharon L. Lohr<sup>1</sup>

### Résumé

Cette note de Sharon L. Lohr présente une discussion de l'article « Théorie et méthodologie des enquêtes par sondage : orientations passées, présentes et futures » où J.N.K. Rao et Wayne A. Fuller partagent leur vision quant à l'évolution de la théorie et de la méthodologie des enquêtes par sondage au cours des 100 dernières années.

**Mots-clés :** Collecte des données; histoire de l'échantillonnage; échantillonnage probabiliste; inférence à partir d'enquêtes.

Rao et Fuller méritent des remerciements pour leur tour d'horizon succinct d'un domaine auquel ils ont tous deux tellement contribué. Ce n'est pas une mince affaire de résumer l'histoire de l'échantillonnage probabiliste et de dégager les futures orientations en 18 pages!

Il est toujours risqué de prédire l'avenir. Cependant, faire l'historique des sondages nous permet de voir comment les pionniers du domaine ont résolu les défis qui se présentaient à leur époque et quel est le lien entre ces défis et leurs solutions et les problèmes qui se posent aujourd'hui.

Pour commencer, comparons les avantages et les inconvénients de l'échantillonnage probabiliste aujourd'hui aux avantages et inconvénients qui étaient perçus au milieu du XX<sup>e</sup> siècle, quand l'usage d'échantillons probabilistes commençait à se généraliser. Les listes qui suivent sont tirées de Parten (1950, chapitre 4); les premiers traités d'échantillonnage publiés par Deming (1950) et par Hansen, Hurwitz et Madow (1953a) donnent des descriptions similaires. Selon Parten, les avantages des échantillons probabilistes, comparativement à la réalisation d'un recensement de la population ou au tirage d'un échantillon de commodité, sont de quatre types :

- A1. Les estimations peuvent être obtenues plus rapidement à partir d'un échantillon que d'un recensement. Moins d'interviews sont nécessaires et, durant les années 1950, le traitement et la totalisation des données pouvaient être effectués plus rapidement pour un petit ensemble de données que pour un grand. Parten a écrit : [*Traduction*] « Cette économie de temps est un avantage particulièrement important dans les études de notre société moderne dynamique. Les conditions évoluent si rapidement qu'à moins de concevoir des méthodes offrant un raccourci pour mesurer les situations sociales, la mesure n'est déjà plus à jour avant que l'enquête ou le sondage soit achevé. » (Parten, 1950, page 109).
- A2. Les estimations à partir d'un échantillon sont moins coûteuses qu'un recensement, parce que moins d'interviews sont nécessaires. Cela se traduit par des coûts plus faibles en dotation et en formation du personnel de terrain.
- A3. L'enquête peut être taillée sur mesure en fonction des estimations d'intérêt. L'échantillonneur peut être plus prudent dans la collecte des données, en posant exactement les questions souhaitées

---

1. Sharon L. Lohr, Arizona State University, Tempe, AZ. Courriel : sharon.lohr@asu.edu.

et en prenant des mesures pour réduire au minimum le biais dû à la non-réponse et à d'autres sources. Au contraire, un recensement peut ne contenir que quelques questions et offrir peu de possibilités de suivi.

- A4. L'échantillonnage probabiliste permet à l'échantillonneur de concevoir l'enquête de manière à obtenir une précision souhaitée et, plus tard, à communiquer la précision réalisée, sans s'appuyer sur des hypothèses de modélisation. Deming (1950, page 10) insistait sur le fait que non seulement les erreurs d'échantillonnage peuvent être calculées à partir d'échantillons probabilistes, mais aussi que les biais de sélection, de non-réponse et d'estimation sont pour ainsi dire éliminés ou contenus dans des limites connues. Hansen et coll. (1953a, page 10) déclaraient : [Traduction] « Les méthodes d'échantillonnage probabiliste permettent d'éviter entièrement de dépendre du jugement pour déterminer la précision. Dans ces circonstances, et avec des échantillons raisonnablement grands, la précision des résultats obtenus à partir de l'échantillon peut être mesurée d'après l'échantillon proprement dit. »

Parten a également examiné les inconvénients du tirage d'un échantillon au lieu de la réalisation d'un recensement :

- D1. Il est difficile de bien faire l'échantillonnage et d'obtenir des échantillons représentatifs. Le fait de ne pas suivre correctement le protocole d'échantillonnage peut introduire des erreurs dans les estimations, et les résultats peuvent être trompeurs si un échantillon est mal conçu ou mal analysé. En outre, étant donné la pénurie de statisticiens d'enquête chevronnés, il est difficile pour la personne qui se propose de réaliser un sondage d'obtenir une aide technique.
- D2. La petite taille de l'échantillon limite l'information que l'on peut obtenir. Dans un échantillon, les observations pour les sous-populations rares sont peu nombreuses. De plus, le nombre de tableaux croisés est limité, parce que le nombre de cas est trop faible pour certaines sous-classifications d'intérêt.

Qu'en est-il de nos jours des avantages et des inconvénients des échantillons probabilistes énumérés par Parten ? Les inconvénients existent toujours. En particulier, la demande d'information plus détaillée, plus à jour et plus complète augmente chaque année (D2). Néanmoins, alors que les enquêtes présentent toujours l'avantage (A3), à savoir qu'elles peuvent être taillées sur mesure pour répondre aux questions d'intérêt, les avantages (A1) et (A2) ont diminué. Durant les années 1950, la collecte de *n'importe quel* type de données était souvent onéreuse. Même les données provenant d'un petit échantillon de commodité pouvaient nécessiter des interviews coûteuses ou une transcription des documents papier demandant beaucoup de travail manuel. En revanche, aujourd'hui, d'énormes échantillons de commodité peuvent souvent être obtenus à un coût beaucoup plus faible, tandis que des échantillons probabilistes, comme ceux de l'*American Community Survey* ou de la *National Crime Victimization Survey*, deviennent de plus en plus coûteux à mesure que les taux de réponse diminuent. L'information provenant d'échantillons de commodité peut également être disponible plus rapidement que les données issues d'un échantillon probabiliste de haute qualité, pour lequel la pondération des données, le calcul des estimations et les vérifications de la qualité prennent des mois.

L'avantage (A4) consistant à pouvoir planifier une précision souhaitée a également diminué. La plupart des grandes enquêtes utilisent encore des méthodes fondées sur le plan de sondage pour communiquer la précision de l'enquête. Toutefois, la marge d'erreur fondée sur le plan de sondage qui est communiquée ne comprend généralement que l'erreur d'échantillonnage et s'appuie sur l'hypothèse implicite que la pondération de l'échantillon a éliminé le biais de non-réponse dans les estimations. À mesure que diminuent les taux de réponse, on se fie de plus en plus au jugement, en utilisant des hypothèses de modélisation, pour déterminer la précision.

Le contexte de la collecte des données est donc différent de ce qu'il était dans les années 1930, les années 1940 et les années 1950, durant lesquelles les techniques d'échantillonnage probabiliste ont été élaborées et mises en œuvre. À l'époque aux États-Unis, l'échantillonnage probabiliste répondait à un besoin urgent de fournir des renseignements plus rapidement et à moindre coût sur l'agriculture, l'activité commerciale, la production manufacturière, les caractéristiques de la main-d'œuvre et d'autres indicateurs sociaux et économiques. Les pionniers des méthodes de sondage ont révolutionné la collecte des données durant cette période. Duncan et Shelton (1978) soutenaient que cette révolution était rendue possible par des progrès parallèles dans les domaines de la théorie statistique, des comptes nationaux du revenu et de la production, de la capacité de calcul et de l'organisation du système statistique.

Bien que les sources de données, l'infrastructure, la technologie et les méthodes disponibles aient changé, le principal problème qui se pose à nouveau aujourd'hui est le même qu'en 1950 : quel est le meilleur moyen de recueillir des données et de faire des inférences à partir de celles-ci pour faciliter la réponse aux questions stratégiques et de recherche ? Si le cadre actuel des échantillons probabilistes n'existait pas, et qu'on nous demandait de construire un système de collecte de données, que ferions-nous ? Pour de nombreux problèmes, nous voudrions construire un système de collecte des données modulaire, pouvant s'adapter à de nouvelles sources de données et de nouvelles technologies de collecte des données. La plupart des méthodes que Rao et Fuller ont passées en revue seraient utiles pour ce système, mais une nouvelle infrastructure et de nouvelles méthodes – et peut-être une autre révolution – sont nécessaires.

À titre d'exemple, considérons le *National Automotive Sampling System* des États-Unis (NHTSA, 2017a). Le système est composé de deux enquêtes. La première porte sur un échantillon probabiliste stratifié à plusieurs degrés de 50 000 à 60 000 rapports d'accident par la police (RAP) tiré de l'univers d'environ 6 millions de RAP annuels, où les RAP concernant des accidents graves sont échantillonnés avec des probabilités plus élevées que ceux concernant des collisions ne comprenant que des dommages matériels mineurs. Les éléments de données extraits des RAP échantillonnés sont codés dans la base de données électronique; aucune information externe aux RAP n'est obtenue. La deuxième enquête est réalisée sur un plus petit échantillon probabiliste d'environ 5 000 RAP avec une collecte des données beaucoup plus intensive, dans le cadre de laquelle des enquêteurs ayant reçu une formation spéciale visitent le lieu de l'accident, inspectent le ou les véhicules accidentés, obtiennent la permission de consulter les dossiers médicaux, interviewent les témoins, et obtiennent d'autres renseignements au sujet de l'accident. Les données de ces deux enquêtes sont utilisées pour étudier les tendances temporelles des accidents automobiles et les effets des caractéristiques des véhicules sur la sécurité routière (voir, par exemple, NHTSA, 2017b), ainsi que pour produire des milliers de documents de recherche.

Mais supposons qu'on nous demande de concevoir ce système de collecte des données en reprenant tout à zéro. Je tiens à souligner que ces suggestions émanent purement de mon imagination et n'ont aucun lien avec quelque projet que ce soit concernant les enquêtes, qui, en raison de considérations pratiques et contraintes budgétaires courantes, doivent avoir une structure d'échantillonnage à plusieurs degrés. Si le système actuel n'existait pas, ne serait-il pas souhaitable de concevoir la première enquête comme un recensement des RAP au lieu du tirage d'un échantillon ? Cette tâche ne serait pas forcément facile. Hetzel (1997) a décrit le processus long et laborieux d'établissement du système américain de statistiques de l'État civil, qui requiert la coopération des organismes d'État et des gouvernements locaux, des procédures de collecte des données uniformes, et des études approfondies pour valider l'exactitude et la couverture des dossiers. La réalisation d'un recensement des RAP nécessiterait pareillement d'énormes investissements initiaux pour élaborer l'infrastructure et obtenir la collaboration des États et des services de police. Après cet investissement, toutefois, la collecte des données serait établie et les RAP pourraient être transmis électroniquement à mesure qu'ils sont recueillis ou mis à jour.

Les avantages de l'obtention d'un recensement de RAP plutôt qu'un échantillon seraient nombreux. Ainsi, les statistiques seraient disponibles beaucoup plus rapidement, puisqu'il ne faudrait pas attendre jusqu'à la fin de l'année de collecte des données pour pondérer et publier les données d'enquête : les statistiques pourraient être mises à jour à mesure que les données sont recueillies. Néanmoins, l'avantage le plus important d'un recensement tiendrait à l'information supplémentaire sur les sous-populations. Cela permettrait de mieux suivre les données pour déceler les risques d'accident possibles. Dans un échantillon de 50 000 unités, un véhicule correspondant à une combinaison particulière de marque/modèle/année pourrait n'être représenté que par une poignée d'observations (si tant est qu'il y en ait); la taille de l'échantillon serait beaucoup plus grande dans le cas du recensement. Dans certaines enquêtes, comme le soulignent Rao et Fuller, des méthodes d'estimation sur petits domaines peuvent être utilisées pour modéliser les résultats pour des sous-populations dont les tailles d'échantillon sont petites. Cependant, dans le cas des données sur les accidents, il est fréquent que l'on s'intéresse à une sous-population soupçonnée de représenter un point aberrant, parce que l'on suspecte que le nombre d'accidents pour une certaine marque d'automobile ou une certaine caractéristique de véhicule est plus élevé que ne le prédirait un modèle. Ces points aberrants ne peuvent pas être décelés par un modèle pour petits domaines. Le seul moyen d'obtenir l'information sur les sous-populations éventuellement aberrantes consiste à recueillir plus de données à leur sujet.

Mais si l'on procède à un recensement des RAP, quand les méthodes de recherche par sondage entreraient-elles en jeu ? Il y aurait inévitablement des données manquantes qu'il faudrait examiner et modéliser, et un plan d'échantillonnage à deux phases pourrait être utilisé pour obtenir des renseignements auprès des États ou des services de police non répondants. Cependant, le principal problème de conception de l'enquête serait double : premièrement, l'échantillonnage pourrait être utilisé pour une vérification du recensement des rapports d'accident, et deuxièmement, l'échantillonnage pourrait être nécessaire pour la partie du système correspondant aux enquêtes sur les accidents nécessitant un travail intensif. Le recensement des RAP fournirait une riche base de sondage pour le système d'enquêtes sur les accidents et d'autres enquêtes. L'information de cette riche base de sondage pourrait être exploitée dans la conception de l'échantillon,

éventuellement en faisant appel à l'échantillonnage équilibré ou à un plan d'échantillonnage qui peut être adapté dynamiquement aux besoins de données et à la base de sondage mise à jour continuellement.

Naturellement, même un recensement des RAP pourrait être obsolète ou insuffisant pour répondre aux besoins de données dans l'avenir. À mesure qu'augmente le nombre de véhicules équipés de caméras et de détecteurs, ou à mesure que les véhicules automoteurs et les systèmes de surveillance deviennent plus fréquents, un échantillon ou un recensement de RAP pourrait être remplacé ou complété par des données recueillies passivement. L'usage accru de sources de données passives à grande échelle soulève d'importantes questions au sujet de la protection de la vie privée et de la propriété des données, qui devront faire l'objet de grands débats et de nombreuses études, et ces questions dépassent le cadre de la présente discussion. Cependant, au-delà des questions sociétales quant à l'éthique de la collecte des données, quelles nouvelles méthodes statistiques sont nécessaires pour répondre à la révolution concernant la disponibilité des données ?

Selon moi, dans un avenir proche, la recherche devra porter sur trois grands domaines interdépendants, qui sont apparentés aux problèmes de recherche auxquels ont fait face Parten, Deming et Hansen au milieu du siècle dernier.

- De meilleures mesures de l'incertitude pour les estimations d'enquêtes. Quand Hansen et Hurwitz (1949, page 365) ont parlé de la supériorité des échantillons probabilistes par rapport aux échantillons au jugé, ils ont souligné que le caractère exempt d'hypothèses de l'inférence à partir d'échantillons probabilistes dépend de l'obtention de taux de réponse élevés : [Traduction] « Au *Census Bureau*, il est habituellement supposé que, si l'information requise est obtenue auprès de plus de 95 % des ménages désignés, on peut se sentir à l'aise de supposer que l'échantillon a été tiré conformément à la théorie de l'échantillonnage, même si des hypothèses peuvent être nécessaires pour les 5 % restants. D'aucuns ont constaté que, dans certaines situations, des problèmes surviennent même si l'on ne formule des hypothèses que pour 5 %. » Deming (1950, page 13) a également utilisé 95 % comme borne inférieure pour la validité de l'inférence à partir d'échantillons probabilistes : [Traduction] « Un échantillon qui est probabiliste à 95 % ou 98 % et qui, pour les 5 % ou 2 % restants, est sélectionné au jugé ou ajusté au jugé pour tenir compte des refus, des personnes absentes du domicile, etc., peut encore être un excellent échantillon, quoiqu'il soit important d'étudier les 5 % ou même les 2 % restants aussitôt que possible. »

Au fil des ans, le seuil de 95 % appliqué à l'utilisation de méthodes d'échantillonnage probabiliste pour l'inférence a baissé, au point où, aujourd'hui, les mêmes méthodes de pondération sont appliquées pour un échantillon avec un taux de réponse de 10 % que pour un échantillon avec un taux de réponse de 95 %. À mesure que les taux de réponse ont diminué, des hypothèses de modélisation de plus en plus fortes ont été formulées au sujet des mécanismes de réponse et des populations sous-représentées et non répondantes, mais l'incertitude au sujet de ces hypothèses n'est généralement pas reflétée dans les intervalles de confiance communiqués, ceux-ci étant encore fondés principalement sur l'erreur d'échantillonnage. Lohr et Brick (2017) ont attribué les échecs récents des sondages aux systèmes utilisés pour calculer les intervalles de confiance ou de prédiction a posteriori, et ont soutenu que de nouvelles méthodes statistiques, reflétant mieux

l'incertitude au sujet des estimations, sont nécessaires pour communiquer les estimations des intervalles. Comme l'a déclaré Parten (1950, page 403) : [*Traduction*] « Il n'est pas inhabituel de constater que des techniques statistiques très perfectionnées utilisées pour mesurer les erreurs aléatoires dans les données sont si biaisées que tous les moyens de correction connus ne permettraient pas à l'enquêteur de déterminer ce que devraient être les résultats corrects. »

- Combiner de multiples sources de données. Des méthodes telles que le couplage d'enregistrements, les enquêtes à bases de sondage multiples et les modèles hiérarchiques peuvent faciliter la combinaison de données, et peuvent aussi être utilisées pour évaluer la qualité des données provenant de différentes sources. Lohr et Raghunathan (2017) ont examiné les méthodes statistiques pour combiner des données provenant de sources différentes, et soutenu que l'utilisation de sources multiples d'information pourrait également aider à résoudre le problème de l'évaluation et de l'intégration des erreurs systématiques dans les estimations de l'incertitude, quoique les méthodes statistiques examinées ne résolvent pas le problème de l'obtention d'estimations pour des sous-populations qui pourraient être absentes de toutes les sources.

Alors que les méthodes d'échantillonnage probabiliste ont bien servi la société au cours des 70 dernières années, elles pourraient jouer un rôle plus limité à l'avenir, peut-être en étant utilisées dans certaines collectes des données pour valider et vérifier d'autres sources de données, au lieu de servir elles-mêmes de sources principales de données. Hansen, Hurwitz et Pritzker (1953b) voyaient sous cet angle les enquêtes postcensitaires durant lesquelles des efforts intensifs étaient faits en vue d'obtenir l'information la plus exacte possible à partir d'un échantillon aréolaire. Leur texte parle aussi de la nécessité de comparer les coûts de l'obtention de statistiques exactes à l'option d'obtenir de plus grandes tailles d'échantillon ou un plus grand nombre de variables. Selon Hansen, Madow et Tepping (1983), c'est précisément quand des changements sociétaux ont lieu que l'absence de biais garantie par l'échantillonnage probabiliste est essentielle et que l'usage ciblé d'échantillons probabilistes de haute qualité peut fournir des évaluations de la couverture et de l'exactitude de l'information provenant d'autres sources de données.

- Enfin, il faut se concentrer à nouveau sur la conception de l'enquête, qui était le sujet des études publiées entre 1920 et 1960. Un système de collecte des données modulaire, s'appuyant sur différentes sources de données pour répondre à différents besoins d'information, pourrait s'adapter à l'évolution des besoins sociétaux et aux nouvelles technologies de collecte de données. Nous avons besoin de plans d'enquête qui sont robustes aux erreurs dans les sources individuelles et permettent d'évaluer ces erreurs, et qui sont également robustes aux changements possibles dans les sources de données.

Comme l'ont souligné Rao et Fuller, le monde de la recherche statistique a répondu à maintes reprises aux besoins d'information de la société grâce à de nouvelles innovations. Les défis que pose le traitement des nouvelles sources de données et des données manquantes sont importants, au même titre que l'étaient les problèmes passés dont la résolution a mené à l'élaboration de l'échantillonnage probabiliste, l'estimation



sur petits domaines, l'estimation de la variance par rééchantillonnage et la théorie de l'imputation. La prochaine révolution en échantillonnage pourrait se profiler juste au tournant.

## Remerciements

Une partie de la présente discussion a pour source une conférence magistrale de l'auteure intitulée « The Essential Survey Statistician » donnée en 2016 dans le cadre du JPSM et qui peut être consultée à l'adresse <https://www.jpsmclasses.umd.edu/Mediasite/Catalog/catalogs/default>.

## Bibliographie

- Deming, W.E. (1950). *Some Theory of Sampling*. New York: Dover.
- Duncan, J.W., et Shelton, W.C. (1978). *Revolution in United States Government Statistics 1926-1976*. Washington, D.C.: U.S. Department of Commerce.
- Hansen, M.H., et Hurwitz, W.N. (1949). Dependable samples for market surveys. *Journal of Marketing*, 14, 363-372.
- Hansen, M.H., Hurwitz, W.N. et Madow, W.G. (1953a). *Sample Survey Methods and Theory. Volume I: Methods and Applications*. New York: John Wiley & Sons, Inc.
- Hansen, M.H., Hurwitz, W.N. et Pritzker, L. (1953b). The accuracy of census results. *American Sociological Review*, 18, 416-423.
- Hansen, M.H., Madow, W.G. et Tepping, B.J. (1983). An evaluation of model-dependent and probability-sampling inferences in sample surveys. *Journal of the American Statistical Association*, 384, 776-793.
- Hetzel, A.M. (1997). *U.S. Vital Statistics System: Major Activities and Developments, 1950-95*. Hyattsville, MD: National Center for Health Statistics. Disponible à <https://www.cdc.gov/nchs/data/misc/usvss.pdf>, dernière consultation le 5 mai 2017.
- Lohr, S.L., et Brick, J.M. (2017). Roosevelt predicted to win: Revisiting the *Literary Digest* poll of 1936. *Statistics, Politics, and Policy*, 8, 65-84.
- Lohr, S.L., et Raghunathan, T.E. (2017). Combining survey data with other data sources. *Statistical Science*, 32, 293-312.
- National Highway Transportation Safety Administration (NHTSA, 2017a). *National Automotive Sampling System (NASS)*. Disponible à <https://www.nhtsa.gov/research-data/national-automotive-sampling-system-nass>, dernière consultation le 5 mai 2017.
- National Highway Transportation Safety Administration (NHTSA, 2017b). *Traffic Safety Facts, 2015*. Disponible à <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812384>, dernière consultation le 17 mai 2017.
- Parten, M. (1950). *Surveys, Polls, and Samples*. New York: Harper & Brothers.