Composante du produit nº 12-001-X
Division des méthodes d'enquêtes auprès des entreprises

Article

Enquêtes auprès des établissements fondées sur un échantillon représentatif de ménages : l'estimateur de Horvitz-Thompson

par Monroe Sirken et Iris Shimizu

Décembre 1999

Statistique Canada Statistics Canada

Canada

Techniques d'enquête, décembre 1999
Vol. 25, N° 2, pp. 213-218
Statistique Canada, N° 12-001 au catalogue

Enquêtes auprès des établissements fondées sur un échantillon représentatif de ménages : l'estimateur de Horvitz-Thompson

Monroe Sirken et Iris Shimizu [1]

Résumé

La Population Based Establishment Survey (PBES) est une enquête sur échantillons liés de ménages et d'établissements dans le cadre de laquelle les listes d'établissements qui effectuent des transactions avec les ménages formant l'échantillon d'une enquête auprès des membres de la population servent de base de sondage pour les enquêtes auprès des établissements. Le présent article décrit l'estimateur de Horvitz-Thompson pour X, soit la somme des valeurs d'une variable pour les transactions de l'ensemble des établissements, applicable à la PBES.

Mots clés : Échantillonnage superposé; transactions des établissements; plan de sondage intégré.

1. Introduction

Quand il n'existe aucune base de sondage indépendante ou que les bases de sondage existantes ne couvrent pas bien les établissements ou ne fournissent pas une bonne mesure de la taille de ces établissements, la Population Based Establishment Survey (PBES) offre un plan de sondage intéressant pour remplacer celui de l'enquête classique auprès d'un échantillon d'établissements. Pareillement, si la variable étudiée concerne une population rare ou imprécise, difficile à observer directement, la PBES offre un plan de sondage intéressant pour remplacer l'enquête classique auprès d'un échantillon de membres de la population.

Le présent article décrit l'estimateur de Horvitz-Thompson applicable à la PBES pour X, la somme des valeurs d'une variable sur les M transactions de R établissements. Représentons par M_j le nombre total de transactions de l'établissement $E_j (j = 1, ..., R)$ durant une année civile particulière. La tâche consiste à concevoir une enquête polyvalente auprès d'un échantillon d'établissements pour estimer les X pour un grand nombre de variables différentes. Habituellement, le plan de sondage des enquêtes auprès des établissements visant à estimer X repose sur l'échantillonnage à deux degrés où les établissements sont sélectionnés avec probabilité proportionnelle à la taille et où les transactions sont les unités sélectionnées au deuxième degré. Les enquêtes auprès des établissements conçues de cette façon exigent des bases de sondage indépendantes assurant une bonne couverture des R établissements et une bonne mesure de la taille des établissements, c'est-à-dire les M_j.

Alors que les listes de ménages et de personnes dénombrées lors des enquêtes auprès d'un échantillon représentatif de membres de la population servent souvent de base de sondage pour d'autres enquêtes du même type (Mathiowetz 1987; Cox, Folsom et Virage 1987), les listes d'établissements qui effectuent des transactions avec les ménages sélectionnés pour l'enquête auprès d'un échantillon représentatif de la population servent rarement de base de sondage pour les enquêtes par sondage auprès des établissements. L'indice des prix à la consommation (IPC), qui dépend de données recueillies auprès des ménages et auprès des établissements (Leaver et Valliant 1995) est une exception notable. Les ménages dénombrés dans le cadre de la Continuing Point of Purchase Survey (CPOPS), enquête menée auprès d'un échantillon représentatif de la population pour établir l'IPC, déclarent les établissements avec lesquels ils ont réalisés des transactions (achat de marchandises). La liste des établissements déclarés dans le cadre de la CPOPS sert de base de sondage pour la Pricing Survey, enquête menée en vue d'établir l'IPC auprès d'un échantillon d'établissements spécialisés dans la vente au détail pour recueillir des données sur les prix d'un panier de biens de consommation.

Il y a plusieurs années, un groupe d'experts du Committee on National Statistics, National Research Council (Wunderlich 1992), a proposé que le National Center for Health Statistics (NCHS) étudie la possibilité d'utiliser les listes de fournisseurs de services médicaux déclarés par les ménages dans le cadre de la National Health Interview Survey (NHIS) comme base de sondage pour les enquêtes nationales auprès des fournisseurs de services médicaux du NCHS. Ces enquêtes étaient, et continuent d'être, conçues de façon indépendante comme des enquêtes classiques auprès d'échantillons d'établissements. [La NHIS est une enquête-ménage permanente réalisée annuellement par le NCHS auprès d'environ 42 000 ménages pour recueillir des statistiques nationales sur la santé des membres de la population civile américaine non placés en établissement (Massey, Moore, Parsons et Tadros 1989)].

1. Monroe Sirken et Iris Shimizu, National Center for Health Statistics, 6525 Belcrest Road, Room 700, Hyattsville, MD 20782, États-Unis. Courriel : mgs2@cdc.gov.

Le NCHS a lancé un programme de recherche sur la PBES pour répondre à la proposition du comité.

Judkins, Berk, Edwards, Mohr, Stewart et Waksberg (1995) ont comparé les caractéristiques opérationnelles et le plan de sondage des enquêtes sur les services de santé associées à la NHIS, d'une part, et de celles dont le plan de sondage est établi de façon indépendante, d'autre part. Judkins, Marker, Waksberg, Botman et Massey (1999) ont procédé à une comparaison grossière du compromis coût-erreur d'une enquête sur les soins dentaires à plan de sondage indépendant et d'une enquête sur les soins dentaires couplée à la NHIS. Ils concluent provisoirement que si l'on peut établir une liste acceptable des établissements et obtenir une mesure raisonnable de leur taille, il est sans doute préférable de réaliser une enquête indépendante. Sinon, on devrait envisager le couplage de l'enquête à la NHIS.

Récemment, les travaux concernant la PBES ont pris une orientation théorique et ont visé à construire pour ce plan de sondage d'autres estimateurs non biaisés nécessitant des données différentes et à développer des formules analytiques pour le calcul de la variance. Les difficultés conceptuelles qui se sont posées au départ ont pu être surmontées une fois qu'il a été reconnu que la PBES est une enquête sur échantillons représentatifs superposés (Sirken 1970). En appliquant la théorie de l'échantillonnage superposé, Sirken, Shimizu, et Judkins (1995), et Shimizu et Sirken (1998) ont obtenu deux versions de l'estimateur de multiplicité non biaisé applicables à la PBES et ont calculé leur variance. Nous présentons ici l'estimateur de Horvitz-Thompson non biaisé applicable à la PBES et sa variance. Les estimateurs applicables à la PBES sont essentiellement une extension à l'échantillonnage à plusieurs degrés, dans des conditions particulières, des estimateurs fondés sur l'échantillonnage superposé à un degré proposé au départ par Birnbaum et Sirken (1965), et décrit par Thompson (1992).

2. Notation

Représentons par M_j le nombre de transactions de l'établissement E_j ($j = 1, ..., R$). Alors,

$$M = \sum_{j=1}^{R} M_j = \text{nombre total de transaction des } R \text{ établissements.} \quad (1)$$

Supposons que N_j = le nombre de ménages effectuant des transactions avec l'établissement E_j ($j = 1, ..., R$), N_{jl} = le nombre de ménages effectuant des transactions à la fois avec les établissements E_j et E_l ($j \neq l$), et N_0 = le nombre de ménages ne réalisant aucune transaction avec aucun établissement. Alors,

$$N^* = \sum_{j=1}^{R} N_j - \sum \sum_{j \neq 1} N_{jl} = \text{nombre total de ménages effectuant des transactions avec les } R \text{ établissements,} \quad (2)$$

et

$$N = N^* + N_0 = \text{nombre total de nénages.} \quad (3)$$

Représentons par X_{jk} la valeur de la variable pour la $k^{\text{ième}}$ ($k = 1, ..., M_j$) transaction de l'établissement E_j ($j = 1, ..., R$). Alors,

$$X_j = \sum_{k=1}^{M_j} X_{jk} = \text{somme des valeurs de la variable sur les } M_j \text{ transactions de établissement } E_j, \quad (4)$$

et

$$X = \sum_{j=1}^{R} X_j = \text{somme des valeurs de la variable sur les } M \text{ transactions du } R \text{ établissements.} \quad (5)$$

3. Le modèle d'erreur d'échantillonnage superposé

On réalise une PBES pour estimer X. Pour commencer, on exécute une enquête par sondage auprès d'un échantillon aléatoire représentatif de n ménages H_i ($i = 1, ..., n$) dans le cadre de laquelle les ménages échantillonnés identifient chacun des établissements avec lesquels ils effectuent des transactions durant une période précises de l'année civile. Après suppression des doubles comptes d'établissement, on réalise une enquête de suivi auprès des r établissements distincts déclarés par n ménages sélectionnés pour l'enquête auprès de l'échantillon représentatif de ménages et chaque établissement E_j ($j = 1, ..., r$) sélectionné déclare indépendamment les variables d'un échantillon aléatoire m_j de ses M_j transactions.

Judkins et coll. (1999) considèrent la PBES comme une enquête par échantillonnage à deux degrés auprès des établissements dans le cadre de laquelle les r établissements qui ont effectué des transactions avec n ménages échantillonnés dans le cadre de l'enquête auprès des ménages sont les unités sélectionnées au premier degré et les m_j transactions ($j = 1, ..., r$) sélectionnées par chacun des r établissements, les unités d'échantillonnage de deuxième degré. Cependant, les caractéristiques du plan de sondage de la PBES deviennent plus transparentes et les estimateurs de la

PBES, ainsi que leur variance, plus faciles à manipuler si l'on modélise la PBES comme une enquête par échantillonnage superposé à deux degrés des membres de la population. Dans la perspective de l'échantillonnage superposé, les ménages sont les unités d'échantillonnage de premier degré et les transactions dénombrables au niveau des ménages échantillonnés, conformément aux règles de dénombrement de la PBES, sont les unités de deuxième degré.

La règle de dénombrement de la PBES précise que chaque ménage faisant partie du réseau de N_j ménages qui ont engagé des transactions avec $E_j\,(j = 1, ..., R)$ est lié au même échantillon aléatoire de taille fixe m_j des M_j transactions de l'établissement E_j. La règle de dénombrement de la PBES sous-entend que les même m_j transactions de $E_j\,(j = 1, ..., R)$ sont dénombrables dans le cadre de l'enquête-ménage auprès de chaque ménage échantillonné appartenant au réseau de N_j ménages qui ont effectué des transactions avec E_j. Du point de vue de l'échantillonnage superposé, les établissements qui effectuent des transactions avec les ménages sont des répondants par procuration pour les transactions qui sont imputables aux ménages. Les ménages qui participent à la PBES ne déclarent pas leurs propres transactions ni les transactions imputables à leurs destinataires en ce qui concerne la règle de dénombrement de la PBES. Les ménages identifient les établissements avec lesquels ils effectuent des transactions et ces établissements sélectionnent les sous-échantillons de leurs transactions qui sont imputables aux ménages échantillonnés et déclarent les variables pour les transactions sélectionnées.

La règle de dénombrement de la PBES produit une configuration des transactions entre les établissements et les ménages qui répartit les N ménages entre R réseaux d'établissement, $A_j\,(j = 1, ..., R)$, où le réseau A_j contient l'ensemble des N_j ménages et est couplé aux M_j transactions de E_j. Bien qu'un même ménage puisse appartenir à plusieurs réseaux, chacune des M transactions est liée de façon unique à un seul réseau.

Le dénombrement des réseaux varie selon que l'on applique les estimateurs de multiplicité de la PBES ou l'estimateur de Horvitz-Thompson. Les estimateurs de multiplicité dénombrent les M_j transactions liées au réseau $A_j\,(j = 1, ..., R)$ chaque fois que les ménages appartenant à ce réseau sont sélectionnés dans l'échantillon de l'enquête auprès des ménages. L'estimateur de Horvitz-Thompson ne dépend pas du nombre de fois que les ménages appartenant aux mêmes réseaux sont sélectionnés dans l'échantillon de ménages. L'estimateur de Horvitz-Thompson applicable à la PBES ne dénombre chaque réseau qu'une seule fois.

4. L'estimateur de Horvitz-Thompson applicable à la PBES

Pour un échantillon de n ménages sélectionnés par échantillonnage aléatoire simple, et un échantillon total de

$$m = \sum_{j=1}^{r} m_j = \text{transactions}, \tag{6}$$

où les sous-échantillons de transactions $m_j\,(j = 1, ..., r)$ sont sélectionnés indépendamment et par échantillonnage aléatoire simple, l'estimateur de Horvitz-Thompson de X applicable à la PBES est

$$X' = \sum_{j=1}^{R} \frac{\alpha_j}{p_j} X'_j. \tag{7}$$

Ici, α_j est une variable aléatoire qui est égale à 1 si n'importe lequel des n ménages échantillonnés appartient au réseau A_j et α_j est égal à 0, autrement, et

$$X'_j = M_j \sum_{k=1}^{m_j} \frac{X_{jk}}{m_j}$$

est l'estimateur non biaisé de $X_j\,(j = 1, ..., R)$ (8)

et

$$p_j = E(\alpha_j) =$$

la probabilité que n'importe lequel des n ménages échantillonnés appartienne au réseau $A_j\,(j = 1, ..., R)$. (9)

X' est un estimateur non biaisé de X si chacun des R établissements effectue des transactions avec au moins un ménage.

Soit

$$q_j = 1 - p_j =$$

la probabilité qu'aucun des n ménages échantillonnés n'appartienne au réseau A_j. (10)

Si n ménages sont sélectionnés par échantillonnage aléatoire simple sans remise,

$$q_j = \frac{\binom{N - N_j}{n}}{\binom{N}{n}}. \tag{11}$$

Si n ménages sont sélectionnés par échantillonnage aléatoire simple avec remise,

$$q_j = \frac{(N - N_j)^n}{N^n}. \tag{12}$$

La mesure des $q_j\,(j = 1, ..., r)$ peut poser deux problèmes. Premièrement, ils dépendent des $N_j\,(j = 1, ..., r)$, quantités qui sont souvent difficiles à déterminer dans le cas

d'enquêtes auprès des établissements. Deuxièmement, il serait difficile de calculer les q_j pour la plupart des enquêtes auprès des membres de la population qui, comme la NHIS, sont fondées sur un plan de sondage complexe.

5. La variance de l'estimateur de Horvitz-Thompson applicable à la PBES

La variance de l'estimateur de Horvitz-Thompson de X peut s'écrire

$$\mathrm{Var}(X') = \mathrm{Var}\,E(X'|\Omega) + E(\mathrm{Var}\,X'|\Omega) \quad (13)$$

où $(X'|\Omega)$ représente la valeur de X' à condition que Ω soit un échantillon fixe de n ménages.

Considérons le premier terme du deuxième membre de (9),

$$\begin{aligned}\mathrm{Var}\,E(X'|\Omega) &= \mathrm{Var}\left(\sum_{j=1}^{R}\frac{\alpha_j X_j}{p_j}\right)\\ &= \sum_{j=1}^{R}\frac{X_j^2}{p_j^2}\mathrm{Var}(\alpha_j)\\ &\quad + \sum_{j=1}^{R}\sum_{l\neq j}\frac{X_j}{p_j}\frac{X_l}{p_l}\mathrm{Cov}(\alpha_j\alpha_l). \quad (14)\end{aligned}$$

Puisque α_j est une variable aléatoire binomiale,

$$\mathrm{Var}(\alpha_j) = p_j - p_j^2 \quad (15)$$

et

$$\mathrm{Cov}(\alpha_j\alpha_l) = p_{jl} - p_j p_l \quad (16)$$

où

$$p_{jl} = 1 - q_j - q_l + q_{jl}^*\,(j \neq l) \quad (17)$$

est la probabilité conjointe que n'importe lequel des n ménages échantillonnés appartienne aux réseaux A_j et A_l, et $q_{jl}^*(j \neq l)$ est la probabilité que les n ménages échantillonnés ne soient liés ni au réseau A_j ni au A_l.

Si n ménages sont sélectionnés par échantillonnage aléatoire simple avec remise,

$$q_{jl}^* = \frac{(N - N_j - N_l + N_{jl})^n}{N^n}, \quad (18)$$

et si n ménages sont sélectionnés par échantillonnage aléatoire simple sans remise,

$$q_{jl}^* = \frac{\binom{N - N_j - N_l + N_{jl}}{n}}{\binom{N}{n}}. \quad (19)$$

Considérons le deuxième terme du deuxième membre de (13),

$$\begin{aligned}E(\mathrm{Var}\,X'|\Omega) &= E\left[\sum_{j=1}^{R}\frac{\alpha_j}{p_j^2}M_j^2\,\mathrm{Var}(\bar{X}_j')\right]\\ &= \sum_{j=1}^{R}M_j^2\frac{\mathrm{Var}(\bar{X}_j')}{p_j}. \quad (20)\end{aligned}$$

$$\mathrm{Var}(\bar{X}_j') = \frac{M_j - m_j}{m_j M_j}\,\sigma^2(X_j) \quad (21)$$

où la variance de population est

$$\sigma^2(X_j) = \frac{\sum_{k=1}^{M_j}(X_{jk} - \bar{X}_j)^2}{M_j - 1}. \quad (22)$$

Supposons que la PBES soit un échantillon auto pondéré. Alors $rp_j(m_j/M_j) = f$, ou f est la fraction d'échantillonnage globale avec laquelle une transaction est sélectionnée. Pour une certaine valeur de f, la taille de l'échantillon des transactions selectionnées dans l'établissement E_j est

$$m_j = \frac{fM_j}{rP_j} = m\frac{M_j/p_j}{\sum_{j=1}^{r}M_j/M_j/p_j}\;j = 1, ..., r. \quad (23)$$

Si nous combinons (14) et (20), nous obtenons la variance de l'estimateur de Horvitz-Thompson de X applicable à la PBES, soit

$$\begin{aligned}\mathrm{Var}(X') &= \sum_{j=1}^{R}\frac{1-p_j}{p_j}X_j^2 + \sum_{j=1}^{R}\sum_{l\neq j}\frac{p_{jl} - p_j p_l}{p_j p_l}X_j X_l\\ &\quad + \sum_{j=1}^{R}\frac{M_j^2}{p_j}\frac{M_j - m_j}{m_j M_j}\sigma^2(X_j). \quad (24)\end{aligned}$$

Les deux premiers termes du deuxième membre de (24) représentent la composante « entre établissements » de la variance due à l'échantillonnage des ménages. Le deuxième terme du deuxième membre disparaît si aucun des N ménages n'effectue de transactions avec plus d'un établissement. Le troisième terme du deuxième membre de (24), qui représente la composante « à l'intérieur des établissements » de la variance due au sous-échantillonnage des transactions, disparaît dans le cas de l'échantillonnage à un degré quand les établissements échantillonnés déclarent les variables pour toutes les transactions. L'échantillonnage à

un seul degré est le choix le plus probable si l'on réalise une PBES à objectif unique plutôt qu'une PBES à objectifs multiples, particulièrement quand la variable étudiée représente un événement assez rare.

6. Conclusion

Qu'il s'agisse de l'estimateur de Horvitz-Thompson proposé ici ou des estimateurs de multiplicité proposés par Sirken, Shimizu et Judkins (1995) et par Shimizu et Sirken (1998), les estimateurs non biaisés applicables à la PBES dépendent de paramètres de multiplicité pour tenir compte de la variation des probabilités de sélection des établissements déclarés lors de l'enquête auprès de l'échantillon de ménages. Cependant, les estimateurs de multiplicité et l'estimateur de Horvitz-Thompson diffèrent en ce qui a trait à la définition des multiplicités et à la probabilité de réussir à recueillir ces renseignements lors de l'enquête de suivi auprès des établissements qui ont été déclarés lors de l'enquête auprès des ménages.

Le fait que les établissements puissent fournir les renseignements sur la multiplicité et la facilité avec laquelle ils peuvent le faire représente un élément clé du choix de l'estimateur applicable à la PBES, si tant est qu'il y en ait, le plus approprié à une application donnée. Les N_j et les $M_j\,(j = 1, ..., r)$ sont, respectivement, les multiplicités que nécessitent l'estimateur de Horvitz-Thompson applicable à la PBES et les estimateurs de multiplicité applicables à la PBES, où N_j est le nombre de ménages réalisant des transactions avec l'établissement E_j et M_j est le nombre total de transactions effectuées avec l'établissement E_j. Il est peu probable que les valeurs des N_j puissent être obtenues facilement, sauf auprès d'établissements, comme les organisations veillant au maintien de la santé, les entreprises de service public ou les compagnies offrant des polices d'assurance propriétaires occupants, pour lesquels les ménages sont les unités de transaction. Par contre, il est vraisemblable que les valeurs des M_j puissent être fournies facilement par beaucoup d'établissements qui ont tendance à enregistrer le nombre total de services fournis, même s'il est peu probable qu'ils connaissent le nombre de ménages auxquels ces services sont prodigués.

La PBES est une option de plan de sondage dont les applications sont nombreuses. Elle permet de coupler les enquêtes par sondage auprès des membres de la population aux fichiers de données des établissements. Comme le mécanisme ne nécessite la divulgation d'aucun identificateur personnel, la PBES n'est pas subordonnée aux préoccupations concernant la confidentialité des renseignements obtenus qui limitent ordinairement l'accès aux fichiers de données des établissements. La PBES permet d'estimer le nombre de transactions réalisées par les établissements dans des circonstances dépassant le cadre des enquêtes classiques auprès d'échantillons d'établissements quand il n'existe pas de bases de sondage indépendantes sur les établissements ou qu'elles sont inadéquates, et dans des circonstances dépassant le cadre des enquêtes classiques auprès d'échantillon de membres de la population quand les variables étudiées ont trait à des populations rares et imprécises qu'il est difficile d'observer directement. Pour déterminer lesquels de ces avantages de la PBES, et peut-être d'autres, sont concrétisables, il faudra comparer les effets des estimateurs de la PBES, d'une part, et des estimateurs des enquêtes classiques auprès des établissements et des membres de la population, d'autre part, sur les coûts et sur l'erreur.

Remerciements

Les auteurs remercient les membres du groupe d'experts sur la National Health Care Survey du Committee on National Statistics d'avoir soulevé le problème étudié dans le présent article, ainsi qu'un examinateur pour ses commentaires très utiles. Les opinions présentées ici sont celles des auteurs et elles ne représentent pas nécessairement les opinions ni les positions officielles du National Center for Health Statistics.

Bibliographie

Birnbaum, Z., et Sirken, M. (1965). Design of sample surveys to estimate the prevalence of rare diseases: Three unbiased estimates. National Center for Health Statistics. *Vital and Health Statistics*, Séries 2, no. 11. Washington, DC : Government Printing Office.

Cox, G.B., Folsom, R.E. et Virage, T.G. (1987). Design alternatives for integrating the National Medical Expenditure Survey with the National Health Interview Survey. National Center for Health Statistics. *Vital and Health Statistics*, Séries 2, no. 101. Washington, DC : Government Printing Office.

Judkins, D., Berk, M., Edwards, S., Mohr, P., Stewart, K. et Waksberg, J. (1995). National Health Care Survey: List Verses Network Sampling. Rapport non publié, National Center for Health Statistics.

Judkins, D., Waksberg, J., Botman, S. et Massey, J. (1999). National Health Interview Survey: Research for the 1995-2004 redesign. National Center for Health Statistics. *Vital and Health Statistics,* Séries 2, no. 126, 76-80. Washington, DC : Government Printing Office.

Leaver, S., et Valliant, R. (1995). Statistical problems in estimating the U.S. consumer price index. Dans *Business Survey Methods,* (Éds., B.G. Cox, D.A. Binder, B.N. Chinnappa, A. Christianson, M.J. Colledge et P.S. Kott). New York : John Wiley & Sons, Inc.

Massey, J.T., Moore, T.F., Parsons, V. et Tadro, W. (1991). Design and estimation for the National Health Interview Survey, 1985-94. National Center for Health Statistics. *Vital and Health Statistics*, Séries 2, no. 110. Washington, DC : Government Printing Office.

Mathiowetz, N. (1987). Linking the National Survey of Family Growth with the National Health Interview Survey: Analysis of field trials. National Center for Health Statistics. *Vital and Health Statistics*, Séries 2, no. 103. Washington, DC : Government Printing Office.

Shimizu, I., et Sirken, M. (1998). More on population based establishment surveys. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 7-12.

Sirken, M., Shimizu, I. et Judkins, D. (1995). The population based establishment surveys. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 1, 470-473.

Sirken, M. (1970). Household surveys with multiplicity. *Journal of the American Statistical Association*, 65, 257-266.

Thompson, S. (1992). *Sampling*. New York : John Wiley & Sons, Inc.

Wunderlich, G.S. (Éd.) (1992). *Toward a National Health Care Survey*: *A Data System for the 21st Century*. Washington, DC : National Academy Press.