

Combining information from multiple surveys to enhance estimation of measures of health[‡]

Nathaniel Schenker^{1,*,*†} and Trivellore E. Raghunathan²

¹*National Center for Health Statistics, Hyattsville, MD, U.S.A.*

²*University of Michigan, Ann Arbor, MI, U.S.A.*

SUMMARY

Survey estimates are often affected by non-sampling errors due to missing data, coverage error, and measurement or response error. Such non-sampling errors can be difficult to assess, and possibly correct for, using information from a single survey. Thus, combining information from multiple surveys can be beneficial. In addition, combining information from multiple surveys can help to reduce sampling error. This article describes four examples of projects undertaken by researchers within and outside the National Center for Health Statistics of the Centers for Disease Control and Prevention, in which information from multiple surveys was combined to adjust for non-sampling errors and thereby enhance estimation of various measures of health. The four projects can be described briefly as follows: (1) combining estimates from a survey of households and a survey of nursing homes to extend coverage; (2) using information from an interview survey to bridge the transition in race reporting in the United States census; (3) combining information from an examination survey and an interview survey to improve on analyses of self-reported data; and (4) combining information from two interview surveys to enhance small-area estimation. The article highlights the goals, techniques, and results from the four projects and discusses issues that can arise when information is combined from multiple surveys. Published in 2007 by John Wiley & Sons, Ltd.

KEY WORDS: complementary surveys; coverage error; measurement error; missing data; non-sampling error; race bridging; self-reported data; small-area estimation

1. INTRODUCTION

Survey estimates are often affected by both sampling and non-sampling errors. Though a particular survey may have been planned to achieve a certain level of sampling error, it is often difficult to

*Correspondence to: Nathaniel Schenker, National Center for Health Statistics, 3311 Toledo Road, Room 3209, Hyattsville, MD 20782, U.S.A.

†E-mail: nschenker@cdc.gov

‡This article is a U.S. Government work and is in the public domain in the U.S.A.

assess non-sampling errors. Sources of such non-sampling errors include: missing data, which can occur, for example, because of non-response; coverage error, which occurs when the population from which the sample is drawn differs from the population of interest; and measurement or response error, which occurs, for example, when survey responses are not completely accurate. Combining information from more than one survey can take advantage of the strengths of the different surveys and can use one survey to supply information that is lacking in another, thereby adjusting for non-sampling errors. Furthermore, the resulting combined estimates may have lower levels of sampling error. Thus, combining information from multiple surveys may provide enhanced estimates of quantities of interest.

This article provides four examples of combining information from multiple surveys to address non-sampling errors due to non-coverage, missing data, and measurement/response error. These projects were conducted by researchers within and outside the National Center for Health Statistics of the Centers for Disease Control and Prevention to enhance estimation of various measures of health. The types of survey errors handled by combining information differ among the four projects. Section 2 highlights the goals, sources of data, techniques, and results from the four projects. Further discussion is given in Section 3, with an emphasis on issues that can arise when information is combined from multiple surveys.

2. FOUR PROJECTS THAT INVOLVE COMBINING INFORMATION

The four projects to be discussed can be described briefly as follows: (1) combining estimates from a survey of households and a survey of nursing homes to extend coverage; (2) using information from an interview survey to bridge the transition in race reporting in the United States census; (3) combining information from an examination survey and an interview survey to improve on analyses of self-reported data; and (4) combining information from two interview surveys to enhance small-area estimation.

2.1. Combining estimates from a survey of households and a survey of nursing homes to extend coverage

Sometimes there is no single survey that has been designed to cover a population of interest, but two or more different surveys have been designed to cover different subsets of the population. In this case, it might be possible to combine information from the different surveys to obtain an improved estimate for the population of interest. This can be thought of as combining surveys to handle coverage error, although in the situation just described, the coverage error is 'by design' in the sense that the individual surveys were intended to cover only the subsets of the population.

Schenker *et al.* [1] combined information for 1985, 1995, and 1997 from the National Health Interview Survey and the National Nursing Home Survey to estimate the prevalence rates of chronic health conditions for the population consisting of residents of both households and nursing homes. The National Health Interview Survey is a household interview survey and is the principal source of information on the health of the civilian non-institutionalized population of the United States. Information about chronic health conditions of household members is obtained by asking one or more respondents from the household about the existence of specific conditions. The National Nursing Home Survey, on the other hand, is a survey of nursing homes, their expenditures, their

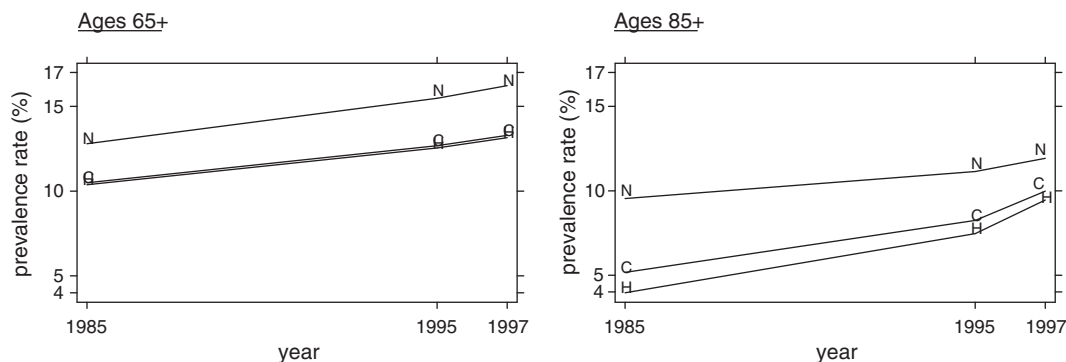


Figure 1. Separate and combined estimated prevalence rates for diabetes, by age group, 1985, 1995, and 1997. C, combined; H, households; N, nursing homes. Source: Schenker *et al.* [1].

current residents, and their staff members, for which data on current residents are obtained *via* interviews with nursing home staff members using information from medical records. Both surveys are conducted by the National Center for Health Statistics. Schenker *et al.* [1] treated the target populations for the two surveys as sampling strata of the combined population and then applied standard design-based techniques for estimating a ratio from complex survey data.

The goal of combining information from the two surveys was to obtain more comprehensive estimates of the prevalence rates of chronic health conditions, especially for the elderly, than would be obtained by using data from either survey alone. It was also desired to avoid misleading results that could occur due to concentrating on a subset of the combined population. Such misleading results could occur, for example, when the prevalence of a condition of interest is much different among household residents than it is among nursing home residents. They could also occur, for example, when changes in the prevalence of a condition over time are of interest, but the propensity of people who have the condition to enter nursing homes is also changing over time.

In 1985, 1995, and 1997, roughly 95 per cent of the combined population of ages 65 and above lived in households, whereas roughly 79 per cent of the combined population of ages 85 and above lived in households [1]. Thus, if the combined population were of interest, excluding nursing homes from analyses would result in only a small amount of non-coverage for the younger elderly (65 and above), whereas it would result in substantially more non-coverage for the older elderly (85 and above).

Schenker *et al.* [1] reported estimates of prevalence rates for several chronic conditions, for the separate household and nursing home populations and for the combined population. Figure 1 displays results for diabetes for illustration. As anticipated above, there is little difference (0.1 percentage point) between the combined estimate and the household estimate for any of the three years for ages 65 and above, whereas the differences for ages 85 and above are substantially larger (1.2, 0.8, and 0.5 percentage points for 1985, 1995, and 1997, respectively).

Similar differences between the age groups occur when comparisons are being made between time points. For example, for ages 65 and above, the increase between 1985 and 1995 is 2.2 percentage points for both the household estimate and the combined estimate. For ages 85 and above, however, the increase is 3.5 percentage points for the household estimate but only 3.1 percentage points for the combined estimate.

2.2. Using information from an interview survey to bridge the transition in race reporting in the United States census

In 1997, the Office of Management and Budget [2] revised the standards for classifying federal data on race and ethnicity. One of the main provisions of the revised standards is that respondents may now choose more than one race group to describe the person in question. In contrast, the previous (1977) standards called for only a single-race group to be reported [3]. As a result of the revision of the standards, when data classified under the 1977 standards are to be compared or combined with data classified under the 1997 standards, there is incomparability. This incomparability can be viewed as a problem of missing data: either the data classified under the 1977 standards are missing their 1997 classifications or *vice versa*.

A particular concern about the incomparability between data classified under the two sets of standards arose at the National Center for Health Statistics in the context of computing vital rates. The data for calculating the numerators of such rates typically come from state-based systems for reporting vital events, and many of these systems still use the 1977 race categories. In contrast, the data for calculating the denominators typically come from the United States census, which began allowing multiple-race reporting in 2000. Overall, counts from the Census 2000 Modified Race Data Summary File indicate that only 1.4 per cent of respondents in the census report multiple races [4], so the extent of the incomparability is limited. For subsets of the United States population, however, the prevalence of multiple-race reporting can be substantially higher. For example, over 35 per cent of the reporting involving American Indians or Alaska Natives is accounted for by multiple-race reports [4]. To enhance comparability between data from the 2000 census and data classified under the 1977 standards, the National Center for Health Statistics, with assistance from the Bureau of the Census, produced estimates of the population counts from the 2000 census that would have been obtained had the 1977 standards been in effect.

The National Health Interview Survey allows multiple-race reporting but also asks multiple-race reporters, in a follow-up question, for the single-race group that best describes the person in question. Bridging models were fitted to data from the National Health Interview Survey for 1997–2000 to predict the single-race group that would have been reported by multiple-race reporters had the 1977 standards been in effect. Proportions generated from the fitted bridging models were then used to distribute multiple-race reporters from the 2000 census into 1977 categories.

The parameter estimates for the bridging models differed among the different multiple-race groups [4–6], suggesting that it is beneficial to fit separate models to the different multiple-race groups to the extent possible. As a consequence of the inclusion of individual-level demographic variables and county-level contextual variables in the bridging models, the percentages of people from a given multiple-race group that were distributed into a given 1977 category varied substantially across the states and the District of Columbia [6], as shown in Figure 2, which displays the approximate distributions of these percentages for the three largest multiple-race groups. At the national level, death rates calculated using the bridged counts from the project were similar to those calculated based on other bridging methods, except in the case of the American Indian or Alaska Native group [6], which had a high involvement in multiple-race reporting as mentioned above.

The ‘bridged’ counts obtained in this project, along with analogous bridged counts for intercensal (between the 1990 and 2000 censuses) and postcensal (after the 2000 census) estimates, have been released publicly (<http://www.cdc.gov/nchs/about/major/dvs/popbridge/popbridge.htm>). Schenker and Parker [5] discussed precursors to the methods used in the project; Ingram *et al.*

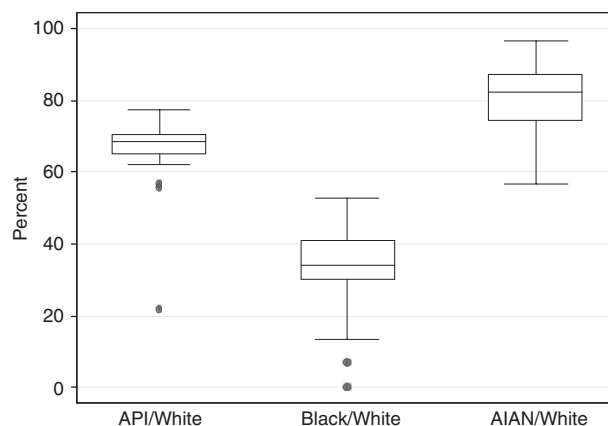


Figure 2. Approximate percentages allocated to White for the three largest multiple-race groups: 50 states and the District of Columbia. AIAN, American Indian or Alaska Native; API, Asian or Pacific Islander. Source: Parker *et al.* [6].

[4] and Parker *et al.* [6] provided descriptions and evaluations of the project; and Schenker [7] developed methods for assessing the variability in estimates due to using bridged counts.

2.3. Combining information from an examination survey and an interview survey to improve on analyses of self-reported data

Self-reported data on health conditions are often based on survey questions of the following form: ‘Has a doctor or other health professional ever told you that you have {the condition in question}?’ Use of such data could lead to inaccuracies in estimating prevalence rates of health conditions. For example, some respondents might not have access to a doctor or other health professional, and thus they might answer ‘no’ even if they have the condition in question. This could be viewed as a type of measurement or response error, in the sense that answers to the self-report question mentioned above might not always accurately measure the clinical existence of a condition. Alternatively, it could be viewed as a type of missing data, in the sense that data on the actual item of interest (clinical diagnosis of the condition) are not available.

Raghunathan [8] and Schenker *et al.* [9] discussed methods for combining information from two surveys conducted by the National Center for Health Statistics to improve on analyses of self-reported data on health conditions. One of the surveys, the National Health and Nutrition Examination Survey, is unusual in that it not only asks self-report questions on health conditions during face-to-face interviews, but it also obtains clinical measures based on physical examinations. The other survey, the National Health Interview Survey, is larger, and it obtains a rich set of variables for use in multivariate analyses, but it relies on self-report questions for its information on health conditions.

‘Measurement error’ models that predict clinical outcomes from self-reported answers and covariates were fitted to data from the National Health and Nutrition Examination Survey, and the fitted models were then applied to data from the National Health Interview Survey to adjust for possible inaccuracies due to self-reporting. Multiple imputation [10] was used to properly reflect the sources of variability in subsequent analyses.

Table I. Unadjusted hypertension prevalence rates, by education level, for people of ages 20 and above, based on data from the 1999–2002 National Health Interview Survey and rates adjusted by using a model fitted to data from the 1999–2002 National Health and Nutrition Examination Survey.

Education level	Prevalence rate (%)	
	Unadjusted	Adjusted
Less than high school	30.9	39.5
High school graduate	22.9	30.1
Some college	18.0	24.5
Four-year college degree	14.5	20.5
Post-graduate degree	17.1	23.6

Table I displays unadjusted and adjusted prevalence rates for hypertension, by level of education, for people of ages 20 and above. The differences between the unadjusted and adjusted rates vary across the levels of education. On an absolute scale, the larger differences tend to occur for those with lower levels of education, whereas on a relative scale, the larger differences tend to occur for those with higher levels of education. Further analyses [8, 9] also suggest that the amount of change due to adjusting for possible inaccuracies due to self-reporting can differ among demographic and socioeconomic subgroups, which might be relevant to studies of health disparities.

2.4. Combining information from two interview surveys to enhance small-area estimation

There is interest among those who study cancer in obtaining small-area estimates of the prevalence of cancer risk factors and cancer screening. For example, such small-area estimates are useful for policy analyses, for examining trends over time, and for use as predictors in modelling future occurrences of cancer. Two major sources of information on cancer risk factors and cancer screening from the Centers for Disease Control and Prevention are the Behavioral Risk Factor Surveillance System, administered by the National Center for Chronic Disease Prevention and Health Promotion, and the National Health Interview Survey, administered by the National Center for Health Statistics.

The Behavioral Risk Factor Surveillance System is a large, state-based, survey that is conducted by telephone. Almost all of the counties in the United States are included in its sample, so it obtains data that can be used to calculate direct estimates at local levels (although many of the county-level estimates are imprecise due to small sample sizes). Because it is a telephone survey, however, it does not include households without telephones in its sample, and it also tends to have relatively high nonresponse rates, as is the case for telephone surveys in general. The National Health Interview Survey is a face-to-face survey that includes both telephone and non-telephone households in its sample, and it has generally lower non-response rates than does the Behavioral Risk Factor Surveillance System. It also has a question that enables analysts to identify each household as having a telephone or not having one. However, it was designed for calculating direct estimates at national and some subnational levels (e.g. for regions and large metropolitan areas) but not at local levels (e.g. for small states or counties). Therefore, it is small relative to the Behavioral Risk Factor Surveillance System, and it includes only about 25 per cent of the counties in the United States in its sample.

In an ongoing joint project involving the National Cancer Institute, the National Center for Health Statistics, the University of Michigan, and the University of Pennsylvania, with assistance from the National Center for Chronic Disease Prevention and Health Promotion, researchers are

Table II. Summaries of BRFSS-alone and BRFSS/NHIS county-level estimates of prevalence rates for current smoking among adult males in 2000, by percentage of households without telephones.

Percentage of households in county without telephones	Mean (standard deviation) of county-level estimates (%)	
	BRFSS-alone	BRFSS/NHIS
< 2	20.6 (3.7)	20.4 (4.4)
2–3	21.1 (3.8)	23.0 (3.7)
3–5	21.9 (4.0)	24.3 (3.9)
5–8	23.0 (4.4)	25.7 (3.9)
8–10	24.1 (4.7)	26.6 (3.8)
10–15	24.4 (4.7)	27.7 (4.1)
15–20	25.4 (4.1)	29.8 (3.8)
≥20	24.1 (5.0)	30.8 (5.8)

Note: BRFSS, Behavioral Risk Factor Surveillance System; NHIS, National Health Interview Survey, The ranges of telephone non-coverage rates are based on the 2000 census. They include the left endpoints and not the right endpoints.

developing methods to combine information from the Behavioral Risk Factor Surveillance System and the National Health Interview Survey to obtain county-level estimates of the prevalence of cancer risk factors and cancer screening [11]. The project aims to reduce sampling error by including information from both surveys and to reduce coverage error as well as possibly errors due to missing data by supplementing the information available from the Behavioral Risk Factor Surveillance System with that available from the National Health Interview Survey.

As an example of the potential differences between telephone and non-telephone households, consider the prevalence rate of current smoking among men in the years 1997–2000. The mean direct estimate across the counties included in the sample for the Behavioral Risk Factor Surveillance System is about 25 per cent. The mean direct estimate across the counties included in the sample for the National Health Interview Survey is also about 25 per cent for telephone households, but it is about 50 per cent for non-telephone households. Table II displays a comparison for the year 2000, by level of telephone non-coverage, of Bayesian small-area estimates of prevalence rates for smoking that combine information from the two surveys, using techniques described in Raghunathan *et al.* [11], with analogous Bayesian small-area estimates based on the Behavioral Risk Factor Surveillance System alone. The combined estimation procedure tends to yield somewhat higher prevalence estimates, especially for counties that have substantial percentages of households without telephones.

3. DISCUSSION

This article concludes with discussions of some issues that can arise when information from multiple surveys is combined and with some concluding remarks.

3.1. Issues that can arise when information is combined from multiple surveys

In this section, issues of incomparability and issues of modelling will be discussed. Note that although the types of problems discussed in this article would typically not be classified as problems of meta-analysis (that is, combining the results of several studies addressing similar

hypotheses), similar issues are often faced in applications of meta-analysis, particularly issues regarding incomparability.

3.1.1. Issues of incomparability. Perhaps the most important issue that arises in combining information from multiple surveys is that of incomparability of the information obtained from the different surveys. All of the references given for descriptions of the projects in Section 2 discuss incomparability to some extent. In particular, Schenker *et al.* [1] contains a section entitled 'Questions to consider when combining estimates,' in which various issues of incomparability are raised. The discussion below focuses on five potential sources of incomparability of information. For each source, an example is drawn from among the various projects described in Section 2.

One potential source of incomparability is differences in the types of respondents and/or the sources of the respondents' information. For example, in the project to combine information from the National Health Interview Survey and the National Nursing Home Survey to extend coverage (Section 2.1), the National Health Interview Survey is based on face-to-face interviews with household members, and the respondents provide much of their information from memory. In contrast, the National Nursing Home Survey is based on face-to-face interviews with nursing home staff, and the respondents consult medical records while providing information.

A second potential source of incomparability is differences in the modes of interviewing. For example, in the project to combine information from the Behavioral Risk Factor Surveillance System and the National Health Interview Survey for small-area estimation (Section 2.4), the former survey uses telephone interviews whereas the latter survey uses face-to-face interviews.

A third potential source of incomparability is differences in the survey contexts. For example, in the project to use information from the National Health Interview Survey to bridge the transition in race reporting in the census (Section 2.2), the census is more publicized and is considered more of a 'national ceremony' than is probably any other periodic government survey (unless, perhaps, national elections are included), so the atmosphere surrounding the census is different from that surrounding the National Health Interview Survey. Moreover, when the bridged census counts are used together with vital statistics data to compute vital rates, it should be noted that information in the census and the National Health Interview Survey are collected in the context of household surveys, whereas information in vital records can be collected in various contexts, such as in a funeral home (for deaths) or in a hospital (for births).

A fourth potential source of incomparability is differences in the sample designs. For example, one consideration in creating multiple imputations is that the features of the sample design should be incorporated into the imputation model [12]. Careful thought needs to be given as to how to accomplish this goal in a project such as that in which models fitted to data from the National Health and Nutrition Examination Survey are applied to data from the National Health Interview Survey to adjust for possible inaccuracies due to self-reporting (Section 2.3), given that the two surveys involved have different sample designs. (See Section 3.1.2 for further discussion of this issue.)

A final potential source of incomparability is differences in survey questions. For example, in the project to combine information from the National Health Interview Survey and the National Nursing Home Survey to extend coverage (Section 2.1), data on the existence of a chronic condition for each household member in the National Health Interview Survey were obtained from answers given by a household respondent to questions (of the form discussed in Section 2.3) about the specific condition. In contrast, data on the existence of chronic conditions for each nursing home

resident in the National Nursing Home Survey were obtained by checking for the condition on a list provided by a nursing home staff member of a primary diagnosis and a limited number of other diagnoses.

3.1.2. Issues of modelling. Often, when combining data from multiple surveys is beneficial, the reason is that data from any single survey are limited in some sense for addressing the analytic problem at hand. In such cases, applying direct estimation techniques to the survey data is often inadequate, and it is necessary to use more complicated modelling techniques. This was the case in the projects discussed in Sections 2.2–2.4. With limited data and complicated modelling, it is necessary to consider many modelling issues, such as what covariates should be included, what forms the models should take, and what models can best deal with small sample sizes.

An example of covariate selection occurred in the project discussed in Section 2.2 (race bridging). The individual-level covariates that were included in the race bridging models were determined in part by which variables were available in the file of census counts to which results from fitting the models were applied. The contextual variables that were included in the models depended in part on what contextual information was available for counties from the 2000 census. Thus, in addition to statistical and subject-matter considerations, operational constraints affected the choice of covariates.

As an example of the form of a model, the issue of differences in the designs of the two sample surveys in the project discussed in Section 2.3 (self-reported data) was dealt with in part by creating classes based on propensity score models for inclusion in one survey *versus* the other. Once these classes were created, the ‘measurement error’ models were fitted separately within the classes.

Finally, as an example of dealing with small sample sizes, in the project discussed in Section 2.4 (small-area estimation), the arcsine square root transformation was applied to the rates that were modelled, in part to simplify the modelling by stabilizing the variances of the outcomes. However, the theory behind the variance stabilizing properties of the arcsine square root transformation is a large-sample theory, and thus the transformation might be less effective for some of the counties in the project that have sparse samples.

3.2. Concluding remarks

Limitations of resources prevent individual surveys from being large enough and extensive enough to address every type of analytic problem that may arise, and it is also not possible to conduct a separate, smaller survey for each such problem. Moreover, the results in projects such as those discussed in Section 2 have been encouraging. Thus, in spite of complicating issues such as those discussed in Section 3.1, combining information from multiple surveys appears to be potentially useful and an important area for further research.

ACKNOWLEDGEMENTS

Raghunathan’s research was partially supported by Centers for Disease Control and Prevention contract 0000359682 and National Institutes of Health grant R24 HD047861. The findings and conclusions in this paper are those of the authors and do not necessarily represent the views of the National Center for Health Statistics, Centers for Disease Control and Prevention.

REFERENCES

1. Schenker N, Gentleman JF, Rose D, Hing E, Shimizu IM. Combining estimates from complementary surveys: a case study using prevalence estimates from national health surveys of households and nursing homes. *Public Health Reports* 2002; **117**:393–407.
2. Office of Management and Budget (US). Revisions to the standards for the classification of Federal data on race and ethnicity. *Federal Register* 1997; **62**:58781–58790.
3. Office of Management and Budget (US). Race and ethnic standards for Federal statistics and administrative reporting. *Statistical Policy Directive No. 15*, 1977.
4. Ingram DD, Parker JD, Schenker N, Weed JA, Hamilton B, Arias E, Madans JH. United States Census 2000 population with bridged race categories. *Vital and Health Statistics* 2003; **2**(135):1–63.
5. Schenker N, Parker JD. From single-race reporting to multiple-race reporting: using imputation methods to bridge the transition. *Statistics in Medicine* 2003; **22**:1571–1587. DOI: 10.1002/sim.1512
6. Parker JD, Schenker N, Ingram DD, Weed JA, Heck KE, Madans JH. Bridging between two standards for collecting information on race and ethnicity: an application to Census 2000 and vital rates. *Public Health Reports* 2004; **119**:192–205.
7. Schenker N. Assessing variability due to race bridging: application to census counts and vital rates for the year 2000. *Journal of the American Statistical Association* 2003; **98**:818–828. DOI: 10.1198/016214503000000756
8. Raghunathan TE. Combining information from multiple surveys for assessing health disparities. *Allgemeines Statistisches Archiv* 2006; **90**:515–526.
9. Schenker N, Raghunathan TE, Bondarenko I. A case study of combining information from an examination-based health survey and an interview-based health survey to improve on analyses of self-reported data. Working Paper Series, Department of Biostatistics, University of Michigan, Ann Arbor, MI, U.S.A., 2006.
10. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Wiley: New York, 1987.
11. Raghunathan TE, Xie D, Schenker N, Parsons VL, Davis WW, Feuer EJ, Dodd KW. Combining information from two surveys to estimate county-level prevalence rates of cancer risk factors and screening. *Journal of the American Statistical Association*, in press.
12. Rubin DB. Multiple imputation after 18+ years (with accompanying articles and discussion). *Journal of the American Statistical Association* 1996; **91**:473–520.