UNDERCOUNT STEERING COMMITTEE (1991). Technical assessment of the accuracy of unadjusted versus adjusted 1990 census counts, June 21, 1991. Bureau of the Census, Washington, D.C.

U.S. DISTRICT COURT (1991). Eastern District of New York (Judge Joseph M. McLaughlin), Memorandum and Order 88 CV 3474, 92 CV 1566, 92 CV 2037, April 13, 1993.

U.S. DISTRICT COURT (1992). *New York et al. v. Commerce Department et al.* U.S. District Court, Eastern District of New York (Judge Joseph M. McLaughlin), 89 CV 3474, 92 CV 1566, 92 CV 2037.

WACHTER, K. W. (member, Special Advisory Panel) (1991). Recommendations on 1990 census adjustment. Report submitted to the Secretary of Commerce, June 17, 1991.

WACHTER, K. W. (1993a). The census adjustment trial: an exchange. *Jurimetrics* **34** 107–115.

WACHTER, K. W. (1993b). Ignoring nonignorable effects. Comment on "Hierarchical logistic regression models for imputation of unresolved enumeration status in undercount estimation" by T. R. Belin, G. J. Diffendal, S. Mack, D. B. Rubin, J. L. Schafer and A. M. Zaslavsky. *J. Amer. Statist. Assoc.* **88** 1161–1163.

WACHTER, K. and FREEDMAN, D. (1994). Local heterogeneity in 1990 census coverage. Technical report, Dept. Statistics, Univ. California, Berkeley.

WACHTER, K. W. and SPEED, T. P. (1991). Comment on "Total error in PES estimates of population: the dress rehearsal of 1988" by M. H. Mulry and B. D. Spencer. *J. Amer. Statist. Assoc.* **86** 858–861.

WEST, K. K. (1991). PES Evaluation Project #P5a: analysis of P-sample fabrication from evaluation followup data, 1990 Coverage Studies and Evaluation Series Memorandum #G-2, Statistical Support Division, Bureau of the Census, Washington, D.C. (July 12, 1991.)

WEST, K., CORBY, C. and VAN NEST, G. (1989). Quality assurance for the clerical matching operation in the 1990 Post-Enumeration Survey. In *Proceedings of the Survey Research Methods Section* 699–703. Amer. Statist. Assoc., Alexandria, VA.

WEST, K. K., MULRY, M. H., PARMER, R. and PETRIK, J. (1991). Address reporting error in the 1990 Post-Enumeration Survey. In *Proceedings on Survey Research Methods Section* 236–241. Amer. Statist. Assoc., Alexandria, VA.

WOLTER, K. (1986a). Comment on "Regression models for adjusting the 1980 census," by D. A. Freedman and W. C. Navidi. *Statist. Sci.* **1** 24–28.

WOLTER, K. (1986b). Some coverage error models for census data. *J. Amer. Statist. Assoc.* **81** 338–346.

WOLTER, K. (1987a). Technical feasibility of adjustment. Memorandum to the Undercount Steering Committee, Bureau of the Census, Washington, D.C.

WOLTER, K. M. (1987b). Comment on "Census undercount adjustment and the quality of geographic population distributions" by A. L. Schirm and S. H. Preston. *J. Amer. Statist. Assoc.* **82** 978–980.

WOLTER, K. M. (1991). Accounting for America's uncounted and miscounted. *Science* **253** 12–15.

WOLTER, K. M. and CAUSEY, B. (1991). Evaluation of procedures for improving population estimates for small areas. *J. Amer. Statist. Assoc.* **86** 278–284.

WOLTMAN, H. F. (1991). Final report for loss function analysis for places. Statistical Support Division, Bureau of the Census, Washington, D.C.

WOLTMAN, H., ALBERTI, N. and MORIARITY, C. (1988). Sample design for the 1990 census Post-Enumeration Survey. In *Proceedings of the Survey Research Methods Section* 529–534. Amer. Statist. Assoc., Alexandria, VA.

WOLTMAN, H. F., GRIFFIN, R. A., NAVARRO, A., BATES, L. and MULRY, M. H. (1991). Post-Enumeration Survey Evaluation Project P16: total error model—loss function evaluation. Statistical Support Division, Bureau of the Census. Washington, D.C.

YLVISAKER, D. (1991). A look back at TARO. Technical report, Dept. Mathematics, UCLA.

ZASLAVSKY, A. M. (1993a). Combining census, dual-system, and evaluation study data to estimate population shares. *J. Amer. Statist. Assoc.* **88** 1092–1105.

ZASLAVSKY, A. M. (1993b). Composite estimators for multiple domains and information sources via hierarchical Bayes linear models and Gibbs sampling. Presentation at UCLA Workshop on Model Selection, Bayes Factors, and Sensitivity Analysis. (February 19, 1993.)

ZASLAVSKY, A. M. and WOLFGANG, G. S. (1993). Triple system modeling of census, Post-Enumeration Survey, and administrative list data. *Journal of Business and Economic Statistics* **11** 279–288.

Comment

Ian Diamond and Chris Skinner

From this side of the Atlantic, the amount of attention devoted to the adjustment of the U.S. census can sometimes seem bewildering. Censuses are conducted for many purposes and raise many methodological problems. For perhaps most purposes, the effects of coverage errors in the census seem likely to be minor compared, for example, to the effects of nonresponse in most sample surveys. Nevertheless, there are important uses of census data in Great Britain, as in the United States, where coverage errors do matter, most notably in the preparation of local area population estimates, which are widely used by central government for the allocation of resources. Furthermore, the adjustment issue has particular significance for the 1991 censuses of England and Wales and of Scotland since, unlike in 1981 and before, the estimates of the national under-

Ian Diamond and Chris Skinner are Professors, Department of Social Statistics, University of Southampton, Southampton S017 1BJ, United Kingdom.

count based on postenumeration data differ considerably from the estimates based on rolling forward the 1981 census to 1991 by adjusting for intercensal births, deaths and net migration.

In our comments, we respond to the Editor's invitation to broaden the discussion by describing some work that has recently been undertaken using data from the 1991 census of England and Wales to obtain population estimates. Following this discussion, we refer specifically to the three papers.

An important aim of the census in England and Wales is to provide a base from which to calculate midyear estimates of the population. For many reasons censuses in England and Wales need to be taken in early spring, typically in April, and so from the outset adjustments have to be built into the estimation process so as to yield midyear estimates. Thus, the issue for the 1991 census was never one of whether to adjust but rather of how to adjust. In addition to the adjustment to the midyear date, there were also initial adjustments made to account for infant undernumeration, and for students and members of the armed forces who were not at their usual address on census night. After these adjustments, the overall undercount in England and Wales has been estimated by two principal strategies. The first has been via a postenumeration survey, the Census Validation Survey (CVS). This gave an estimated undercount of 240,000 residents. The second strategy has been to use demographic methods to roll forward the population count from 1981. This gave an estimated undercount of 1,113,000 residents. After considerable work it was decided that the rolled forward estimate provided the most reliable figure on which to base the overall population estimate. This meant that after adjusting for the CVS there remained a difference of 873,000 residents between the estimates based on the two alternative strategies.

Much work has been undertaken to attempt to explain this discrepancy (e.g., Office of Population Censuses and Surveys, 1993). For ages 45–79, the two sets of estimates are in good agreement. For the under-45's, the decision to believe the estimates from the rolled-forward approach was supported by a comparison of the demographic structure of the population under the two approaches. This showed that the age/sex structure of the census count adjusted by the CVS was markedly short of young people, predominantly male, particularly those in their 20's—groups which are universally acknowledged as hard to enumerate. As a result, the final national count and age/sex distribution was adjusted at ages 0–44 to match the rolled-forward estimates. At ages above 80, auxiliary data from the Department of Social Security on pensions were used to adjust to the final midyear estimate.

The shortfall of the CVS undercount estimate has led to scrutiny of the methodology of coverage error estimation based on postenumeration surveys. The British approach is quite distinct from the dual system approach in the United States. Instead of attempting an independent replication of the census count, "dependent reenumeration" is used to follow up the census in sampled areas with the aim of assessing the "true" count. Experienced interviewers are given data from household census forms and conduct reinterviews to assess the coverage and quality of the responses. In addition, buildings are relisted, nonresidential buildings are checked, addresses recorded in the census as vacant or absent are followed up and the occupation of buildings with multiple households is checked.

The undercount rate can then simply be estimated by comparing the census count with the revised "true" counts in the sample areas. The accuracy of estimates based on such "dependent reenumeration" as compared to dual system estimates clearly depends on the validity of the assumptions underlying each procedure, as discussed by Breiman for the U.S. case. In principle, the dependent reenumeration approach avoids the problems of correlation bias in the dual system estimate and is less critically dependent on the matching of census respondents to postenumeration survey respondents, which as Breiman emphasises may be prone to error. On the other hand, the assumption that the "true count" in sampled areas can be determined exactly is clearly heroic and is strongly refuted by the evidence referred to earlier. A considerable number of persons seem to have been missed on both occasions either because of deliberate avoidance or because of problems in making contact. There is a clear need to develop a new strategy for the CVS for the next census in England and Wales, perhaps adopting a form of dual system estimation. This will require more comprehensive as well as bigger samples.

Although methods of demographic analysis such as those described above have strong advantages for the estimation of coverage error at a national level, problems of estimating internal migration make them difficult to use to distribute the estimated undercount geographically. In England and Wales, infants, armed forces personnel and students can be allocated to local areas using vital registration and administrative data. However, for the majority of the population, local area population estimates were obtained by using a demographic estimation procedure to allocate weights to local area counts.

The local authorities were divided into 10 broad types of areas (based on a classification of 1981 census data). The sex ratios (males per 100 females) obtained by adjusting each five-year age group in each

local area type by the national undercount for that age group were then compared with sex ratios calculated from the 1971 and 1981 estimates. Since several of the 1991 estimates seemed implausible, a procedure for adjusting the estimates to attain a "target" sex ratio in each of the 10 groups of areas was developed (Diamond, 1993). After much exploratory analysis it was decided to use this adjusted approach only in the three five-year age groups 20–24, 25–29 and 30–34. This was for two main reasons. First, there was little evidence of implausible sex ratios as a result of using a uniform adjustment in other age groups. Second, the majority of the estimated national undercount occurred in this age range. Adjustments using the demographic estimation approach were incorporated into the final midyear estimates for local authorities and have been generally accepted although work to estimate populations for smaller areas continues. Whereas synthetic estimation has so far not been used explicitly, some methodological investigation of these techniques has taken place (e.g., Skinner, 1991).

We now turn specifically to the three papers under discussion. Breiman draws our attention to a range of sources of nonsampling error which may affect the dual system estimates. His discussion draws on the extensive evaluation work of the U.S. Census Bureau, and Breiman's attempt to summarise the combined effects of these errors on dual system estimates might be viewed as offering an alternative to the total error analysis reported, for example, in Mulry and Spencer (1991, 1993). We are not qualified to comment on the detailed differences between these alternatives but, like Belin and Rolph, find it hard to view the point estimates of Breiman as a methodological advance over the interval estimates of Mulry and Spencer.

We have described how, in the British case, demographic analysis has been used to validate postenumeration survey methodology at a national level. For the U.S. case, we note that the 95% interval (1.00%, 2.25%) of Mulry and Spencer (1993) for the national undercount rate agrees well with the estimate from demographic analysis of 1.85% (Robinson, Ahmed, Das Gupta and Woodrow, 1993). We would be interested to hear how Breiman views differences between results in his Table 15 (and perhaps similar results disaggregated by age and gender) and estimates from demographic analysis.

Freedman and Wachter examine the impact of the heterogeneity of undercount rates between states within post strata. Belin and Rolph refer to this impact as heterogeneity bias. We feel this risks confusion with the biasing effect on the dual system estimator of heterogeneity of capture probabilities between different individuals (correlation bias). Freedman and Wachter's discussion focuses not on properties of the dual system estimator *per se* but rather on the use of synthetic estimation across states. Like Belin and Rolph and others, we do not find the heterogeneity between states surprising. It seems more natural to interpret the impact of the heterogeneity in terms of variance rather than bias. Thus, the important questions that we feel flow from Freedman and Wachter's work are (a) how well can the additional uncertainty induced by the heterogeneity be quantified in appropriate interval estimators and (b) are there ways that this additional uncertainty can be reduced (for example, the procedures for England and Wales, described earlier, implicitly assume that sex ratios can be used as auxiliary variables at the local area level to reduce error). We encourage Freedman and Wachter to develop work in these directions.

As to the future of adjustment methodology, the experience of England and Wales suggests great scope for demographic estimation, and we are encouraged to see developments in that direction in the United States (Robinson, 1994). This point has been well made in an excellent review by Himes and Clogg (1992). However, if the method used is to be based on an analysis of intercensal birth records (Robinson, 1994), then it will be necessary to assess the accuracy of birth records as suggested by Wachter (1994).

Belin and Rolph provide wise comments on the nature of the adjustment debate. The importance that this debate take place in an open and scientific manner cannot, we feel, be overemphasized. Not only does it allow census agencies in other countries to share experiences, but it is to be hoped that it also enables our understanding of methodology to progress in a cumulative manner. To come to a consensus on adjustment does, of course, require more than just scientific progress. Our experience is that for there to be a common will to adjust there must first be an agreement that certain groups will be harder to enumerate than others and so a disproportionate adjustment for some groups will be seen to be fairer than either a uniform adjustment nationally or even no adjustment at all. In England and Wales the general agreement that young people, particularly men, in inner city areas were more likely to be underenumerated led to an acceptance of the demographic method used to estimate the mid-1991 population. It is also important for there to be a common recognition that the final census figure is an estimate subject to error. Clearly such errors must be minimized and standard errors estimated, but to expect that a census estimate must be perfect is unrealistic.

ACKNOWLEDGMENT

The research for this discussion was supported by ESRC Grant H507255122.