

- of Hutterite Data," in *Sociological Methodology*, 1987, ed. C. Clogg, Washington, DC: American Sociological Association, pp. 247-307.
- Lessler, J. T., and Sirken, M. G. (1985), "Laboratory-Based Research on the Cognitive Aspects of Survey Methodology," *Milbank Memorial Fund Quarterly*, 63, 565-581.
- Sawyer, D. R., and Sawyer, D. R. T. O. (1987), "Malaria on the Amazon Frontier: Economic and Social Aspects of Transmission and Control," technical report, Center for Development and Regional Planning, Federal University of Minas Gerais, Brazil.
- Tourangeau, R. (1984), "Cognitive Science and Survey Methods," in *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines*, eds. T. Jabine, M. Straf, J. Tanur, and R. Tourangeau, Washington: National Academy Press, pp. 73-100.

Comment

JOEL E. COHEN*

1. INTRODUCTION

Cleland and Verma give a clear, balanced appraisal of the World Fertility Survey's methodology. I comment on three issues not considered in their article. My comments look to methodological issues that affect the analysis of WFS data and the collection of future data, and are not intended as criticisms.

First, if fertility and mortality interact, historical inferences based on facts about survivors may be biased by selective effects of mortality. Second, if women are sensitive to questions about the timing and circumstances of past sexual unions, conceptions, and pregnancy outcomes, then randomized responses might be used in future surveys to reduce evasive or false answers (at the cost of increased variance). Third, if retrospective questions deal with matters that are subject to cultural preconceptions or the mood of the respondent at the time of questioning, the answers, even if intended to be honest, are subject to systematic distortions.

2. INTERACTIONS OF FERTILITY AND MORTALITY

Suppose that women aged 15-49 years are interviewed at a time called "now." The 49-year-olds have survived the mortality experienced by the cohort that was age 15 34 years ago, whereas the women of that cohort who died before reaching age 49 are absent from the survey. The survivors are a subset of the original cohort, selected by mortality.

Assume for simplicity that mortality has not changed in the past, and consider correlated heterogeneity in fertility and mortality. To take a simple case, suppose each cohort of women contained two subcohorts, one with high fertility and one with low fertility. Suppose also that the high-fertility subcohort had either high or low mortality, and that the low-fertility subcohort had the opposite mortality. Higher fertility might be associated with higher mortality if more frequent pregnancy weakened a mother's consti-

tution or meant more frequent exposure of the mother to the risks of childbearing. Higher fertility might be associated with lower mortality if basically healthier women conceived and bore live children more frequently.

A cross-sectional retrospective survey of women cannot give information about the strength or direction of the association, if any, between fertility and mothers' mortality. But retrospective inferences may be subject to bias if the association exists.

If higher fertility were associated with higher mortality (both unchanging over time), today's surviving 49-year-olds would contain more survivors with low fertility than with high fertility, relative to the original composition of their cohort. So the fertility rate for the 15-year-olds of 34 years ago, estimated on the basis of the 49-year-olds who survive now, would be lower than the actual fertility rate of the 15-year-olds 34 years ago, and lower than the fertility rate of women 15 years old now. In short, for a given age group, a rising trend over time in the age-specific fertility of 15-year-olds (e.g.) might be inferred as a result of selective removal by mortality of women with high fertility.

On the other hand, if higher fertility were associated with lower mortality (both unchanging over time), the argument would run in the opposite direction. For a given age group, a falling trend over time might be inferred as a result of selective removal by mortality of women with low fertility.

The argument has assumed no changes over time in the age-specific fertility and mortality rates. The calculations become more complex when age-specific fertility and mortality rates change over time, but the main point remains the same. In the presence of heterogeneity, studying only a cross-section of survivors may lead to faulty historical inferences. [For reviews of the sometimes perverse consequences of demographic heterogeneity, see Vaupel and Yashin (1985) and Cohen (1986).]

How serious is this theoretical possibility in the case of

* Joel E. Cohen is Professor of Populations, Rockefeller University, New York, NY 10021. This work was supported in part by U.S. National Science Foundation Grant BSR 87-05047.

WFS data? In response to a previous draft of this comment, John Cleland wrote me (personal communication, August 1987): "Even with high adult mortality and a high correlation between childbearing and survivorship, it can be demonstrated that estimates are little affected (World Fertility Survey Technical Bulletin 4)." Accepting this, I conclude that a future analyst should be alert to the possibility and the magnitude of selective effects, especially when studying a quantity or parameter whose sensitivity to selection has not been investigated in WFS Technical Bulletin 4.

3. SENSITIVE BEHAVIOR AND RANDOMIZED RESPONSES

If women are sensitive to questions about the timing and circumstances of past sexual unions, conceptions, and pregnancy outcomes, then randomized responses might be used in future surveys to reduce evasive or false answers (at the cost of increased variance). I find it implausible that married women in a traditional society would not be sensitive about sexual unions or conceptions that their husbands or mothers-in-law might not know about.

Randomized responses were used effectively as long ago as 1965 (Horvitz, Shah, and Simmons 1967) to find the frequency of illegitimate births, and they were used (Abernathy, Greenberg, and Horvitz 1970; Gould, Shah, and Abernathy 1969; Greenberg, Abernathy, and Horvitz 1970) to study other sensitive aspects of reproductive behavior, such as then-illegal abortions. Fiddler and Kleinknecht (1977) and Dawes and Moore (1980) used randomized responses to investigate the sexual behavior of college students. Dawes and Moore compared the frequency of unconventional sexual and other behavior estimated using randomized response, with the frequency estimated using direct questions with an assurance of confidentiality. In many cases, there were statistically significant differences in the estimated frequencies. Whenever such a difference was found, the higher frequency of unconventional behavior was estimated using randomized response. Fiering and Hooper (1985) analyzed a bivariate form of randomized responses and suggested that it could be used to study risk factors associated with sensitive behavior or conditions.

In response to a previous draft of this comment, John Cleland wrote (personal communication, August 1987):

The randomized response technique was considered briefly at the start of the [WFS] program, but was quickly ruled out on grounds of practicability and acceptability. It is true that WFS data on induced abortion are usually worthless because of obvious underreporting. But for other topics which might be sensitive (e.g., contraception, illegitimate births, marriage dissolution), the data do not appear to have suffered from deliberate concealment. Of course, strict validation was not possible, but internal consistency checks and comparisons with other surveys have usually failed to reveal major defects of this type.

Forms of randomized response questions simple enough to be used and understood in the field remain to be developed. It also remains to be determined whether the cost in increased variance of randomized response questions is compensated by the reward in reduced evasion.

4. PSYCHOLOGICAL INFLUENCES ON ACCURACY OF RECALL

If retrospective questions deal with matters that are subject to cultural preconceptions or the mood of the respondent at the time of questioning, the answers, even if intended to be honest, are subject to predictable and systematic distortions.

Unlike computer memory, human retrospection is largely reconstructed. A person's reconstruction of the past appears to be strongly influenced by his or her present state and his or her theories about the past. Dawes and Pearson (1986) reviewed many striking examples in which individuals were asked to recall their own physically observable behavior. For example, high school students were interviewed about their use of alcohol, tobacco, and marijuana in three waves of a panel study: initially, after one year, and after 2½ years. Among students whose behavior changed from wave to wave, half to three-quarters of the variance in the initial behavior as recalled after 2½ years could be explained entirely by the students' final behavior. The retrospections were least accurate for individuals who changed their behavior during the study, and most accurate for individuals who did not change their behavior.

This and several other examples suggest that individuals or conditions that changed during the period of recall are most vulnerable to biased recall. By analogy, retrospective recall of past use of contraceptives, for example, may be more reliable for individuals who did not change their pattern of use than for individuals who did change their pattern of use.

Single or unique events, Dawes and Pearson found, generally are more accurately recalled than similar, frequently repeated, or bunched events. By analogy, the time of first marriage may be more reliably recalled than the timing or number of individual conceptions if these have been frequent. The timing of a sterilization may be more accurately recalled than the timing of a frequently used contraceptive.

Not surprisingly, Dawes and Pearson also found that the more distant the events to be recalled, and the fewer the opportunities to rehearse recall in the interim, the less likely the recall is to be accurate.

They suggested:

In order to understand and catalogue the type of distortions that can be expected from retrospective recall, it is necessary to conduct studies that are *simultaneously* prospective and retrospective. That requires a substantial longitudinal commitment of research resources. . . . One cannot, however, avoid the conclusion that the resources allocated toward the measurement and understanding of the bias and error variance of retrospectively collected data have been an imprudently small fraction of the total resources devoted to collecting and using such accounts. (pp. 20–21, 22)

Refinements in statistical methodology—sampling and analysis—need to be complemented by improvements in the measuring instruments that generate the data. Enhanced collaboration between survey demographers and cognitive psychologists could benefit both demographers and psychologists.

ADDITIONAL REFERENCES

- Abernathy, J. R., Greenberg, B. G., and Horvitz, D. G. (1970), "Estimates of Induced Abortion in Urban North Carolina," *Demography*, 7, 19-29.
- Cohen, J. E. (1986), "An Uncertainty Principle in Demography and the Unisex Issue," *The American Statistician*, 40, 32-39.
- Dawes, R. M., and Moore, M. (1980), "Die Guttman-Skalierung Orthodoxer und Randomisierter Reaktionen," in *Attitude Measurement*, ed. F. Peterman, Göttingen: Verlag für Psychologie C. J. Hogrefe, pp. 117-133.
- Dawes, R. M., and Pearson, R. W. (1986), "The Effects of Theory-Based Schemas on Retrospective Data," summary of 1986 summer workshop of the Committee on Cognition and Survey Research, Social Science Research Council, New York.
- Fiddler, D. S., and Kleinknecht, R. E. (1977), "Randomized Response Versus Direct Questioning: Two Data-Collection Methods for Sensitive Information," *Psychological Bulletin*, 84, 1045-1049.
- Fiering, M. B., and Hooper, R. (1985), "Analysis of Disclosure Avoidance Procedures," *Civil Engineering Systems*, 2, 12-19.
- Gould, A. L., Shah, B. V., and Abernathy, J. R. (1969), "Unrelated Question Randomized Response Techniques With Two Trials per Respondent," in *Proceedings of the Social Statistics Section, American Statistical Association*, pp. 351-359.
- Greenberg, B. G., Abernathy, J. R., and Horvitz, D. G. (1970), "A New Survey Technique and Its Application in the Field of Public Health," *Milbank Memorial Fund Quarterly*, 48, 39-55.
- Horvitz, D. G., Shah, B. V., and Simmons, W. R. (1967), "The Unrelated Question Randomized Response Model," in *Proceedings of the Social Statistics Section, American Statistical Association*, pp. 65-72.
- Vaupel, J. W., and Yashin, A. I. (1985), "Heterogeneity's Ruses: Some Surprising Effects of Selection on Population Dynamics," *The American Statistician*, 39, 176-185.