

# The Role of Data in Research and Policy

BARBARA A. ANDERSON

## Abstract

Data are essential for scientific research and policy planning. However, there needs to be attention to data quality and to the estimates and models based on those data. In addition, data need to be freely available for researchers to test new ideas and validate the work of others through replication, while respondents who provide data need to be protected. Three issues concerning data are addressed: (i) availability and accuracy of data for new research and reanalysis while protecting human subjects, (ii) problems with the estimation of indicators based on flawed or nongeneralizable data, and (iii) the use of data to develop models for projecting the future, the assumptions on which those models are based, and the assessment of the accuracy of past projections. In each of these areas, increased attention is necessary on how data are used, interpreted, and made available to the scholarly and policy community.

## INTRODUCTION

Data are essential for scientific research and policy planning. In light of this, there has been an increasing call for evidence-based policy in numerous areas. However, there needs to be attention to data quality and to the characteristics of estimates and models based on those data.

Three issues of special concern are as follows:

*Availability and accuracy of data for new research and reanalysis while protecting human subjects.* There are impediments to data sharing for new research and for replicability of findings, and there are inherent tensions between data access and protection of human subjects.

*Problems with the estimation of models and indicators based on flawed or non-generalizable data.* Indicators must be estimated from sufficiently accurate and unbiased data so that a misleading picture of the situation is not given. Methods of estimating indicators are often developed based on relationships in settings where there are high-quality data. The indicator might not have the

same relation to the data in settings in which less is known, but this is very difficult to determine. The effort to use more appropriate data can lead to additional problems.

*The use of data to develop models for projecting the future, the assumptions on which those models are based, and the assessment of the accuracy of past projections.* Models of the trajectory of demographic changes are necessarily constructed based on situations where a particular change has already occurred. Projecting the future requires assumptions that can be arbitrary and can change radically as the situation in demographically advanced countries changes, producing an unstable situation for users of the projections.

In each of these areas, increased attention is necessary on data are used, interpreted, and made available to the scholarly and policy community.

## FOUNDATIONAL RESEARCH

Some of the main principles of data collection and analysis are as follows:

*Data should be available for replicating the work of others and for testing new ideas.* There has been an increasing standard that data must be made available to others in the scientific community for reanalysis, and an increasing number of scholarly journals require that data used in an article be available at an accessible data repository (Anderson, Greene, McCullough, & Vinod, 2008). Data that are collected using U. S. government funds are to be made available to other researchers in some form, usually within 2 years of the completion of data collection.

*The identity, answers, and other data regarding respondents need to be effectively protected.* Concern with protection of human subjects increased in response to Nazi medical experiments as well as disclosure of egregious violation of the rights of patients in the United States. The Tuskegee Study from 1932 through 1972 studied 600 poor African-American men in Tuskegee, Alabama, 399 of whom already had syphilis. Participants were not told they had syphilis and were not treated for syphilis, even after penicillin was accepted as an effective syphilis treatment in 1947. Many participants and their wives and children died of syphilis. In 1972, information about the study was leaked, which led to the study's termination (Fairchild & Bayer, 1999; Jones, 1981). Disgust at the Tuskegee Study motivated the Belmont Report, which is the basis for the Institutional Review Board guidelines that govern all academic research in the United States.

*It is better to know something rather than nothing in a given situation.* Models of behavior are often based on the assembly of all available high-quality data. Then, a model is fit to those data, which then can be used in other situations. If the model is intended for use in estimating indicators for situations in which the most appropriate data are not available, then the input data should be

readily available in those situations in which one wants to produce estimates. There is always tension between including as much data as possible in the development of the model and being sure not to include low-quality data or biased data.

An example of this is the Coale–Demeny regional model life tables (1966). These mortality models were based on 326 life tables that were thought to have high-quality data. Of these, 324 were from Europe or North America. Almost all the life tables with high mortality were from historical Europe. There was awareness that the life tables used might not represent all of world experience, but it was viewed as too risky to include life tables from low-income questions where the data quality might be poor.

*Sometimes, the only guide to the future is what has happened in the past.* Modeling mortality as addressed in the previous point has its pitfalls, but modeling fertility can be even more risky. Family planning programs in Taiwan and elsewhere in much of Asia in the 1960s and 1970s were quite successful. These programs seemed to indicate that once easy-to-use, effective contraceptives were made available, fertility would rapidly fall. That view was called into question when fertility in sub-Saharan Africa and some parts of Latin America seemed much more resistant to change (Caldwell & Caldwell, 1988).

Population projections are important in and of themselves for planning the future and are also part of the input information for many other purposes, such as estimating the adequacy of future food supply in a region. There have been noted instances in the past when projections were very far off, because of an unknown changeable future.

## CUTTING-EDGE RESEARCH

Research has pointed out shortcomings and considerations about data, models, and estimation. Such studies highlight further work that needs to be done.

### AVAILABILITY AND ACCURACY OF DATA FOR NEW RESEARCH AND REANALYSIS WHILE PROTECTING HUMAN SUBJECTS

Data access for original research and for replicability of findings remains an issue in both low-income and high-income countries. In both settings, issues of researcher or institutional hoarding and of protection of respondent confidentiality arise.

In many low-income countries, there remains lack of high-quality demographic data upon which to base population estimates and to look at interrelations between demographic and social variables. To address this problem, 49 demographic surveillance sites have been established in 20 low-income

countries, especially in sub-Saharan Africa. This is called the INDEPTH Network (2014). In these sites, intensive efforts are made to collect information on demographic events at frequent intervals.

There is concern about access to data from demographic surveillance sites. Data collection requires a huge effort. Persons directing and conducting data collection are often reluctant to turn the data over to external researchers who have not devoted the same amount of energy to this effort. On the other hand, scientific standards require that data be available for independent examination and validation. If other researchers do not have access to the data, it is not possible for alternative explanations to be investigated, and it can lead to questioning of the value of research results based on the data. In addition, since data collection is so time intensive, researchers often need to turn their attention to the next round of data collection as soon as one round is completed. Thus, much data from such sites are analyzed only locally to a limited extent. These issues have led to a lively debate about the conditions under which data from a demographic surveillance site should be available to the larger scholarly community (Baiden, Hodgson, & Binka, 2006; Carrel & Rennie, 2008; Chandramohan *et al.*, 2008).

The balance between protecting respondents, giving those who collect data a fair chance to benefit from publication, and allowing replicability remains a source of tension. In the INDEPTH research sites in many low-income countries, there has been a debate concerned with whether in principle, researchers who were not involved in data collection should have access and whether their research aims should be required to be in line with or whether they should be required to collaborate with project researchers. These potential requirements conflict with principles of independent assessment and replicability in science.

There also remain concerns about sharing of data in the United States. Even as more and more journals require deposit of data in an archive, this had not at least in 2009 affected many open-access journals, which are advocates for total openness in research (McCullough, 2009). Even when data have been deposited in an archive, McCullough, McGear, and Harrison (2008) found it was only possible to replicate the work in 14 out of 62 articles.

As more and more analyses for policy purposes are based on surveys and microdata from censuses, there is concern with protecting the identity of individuals. This often leads to masking data in various ways, including perturbation of data by introduction of a random factor, grouping dates of occurrence of events into 5-year or 10-year periods, and masking aggregating geographic location to a fairly large area. These respondent-protection measures can lead to erroneous conclusions, though, in analysis of publicly available data.

A slight adjustment of ages through introduction of a random factor can sometimes lead to anomalous results, such as unreasonable sex ratios (Alexander, Davern, & Stevenson, 2010; U. S. Census Bureau, 2010). Adjustment of the reported data to mask the identity of those over age 65 can also lead to inaccurate estimates of characteristics of the elderly, such as their income (Fisher, 2010). On event history analysis, the detailed dating of events and the sequencing of events are important, which are not possible with aggregated times of occurrence (Freedman, Thornton, Camburn, Alwin, & Young-DeMarco, 1988), a common way of masking identity through grouping time into 5- or 10-year periods, and reporting data for fairly large geographic units calls for rethinking of how respondent identity should be masked.

Masking geographic detail can help protect respondents (Sherman & Fetters, 2007). However, researchers have increasingly incorporated detailed information about the characteristics of small geographic areas in order to identify clusters of people with particular diseases or who are studying attitudes or behavior need very detailed geographic information to do so (Berg, Stewart, Stewart, & Simons, 2013; Armstrong, Rushton and Zimmerman, 1999). A researcher can apply to the body that controls the data and ask for more detailed information. If the controlling body sees the proposed research as sufficiently valuable, the researcher could obtain the more detailed data, but the approval process can take a long time and is often not successful.

#### PROBLEMS WITH THE ESTIMATION OF INDICATORS BASED ON NONGENERALIZABLE OR FLAWED DATA

Estimates of the number of persons with a disease are sometimes based on results of a survey. In order for the estimates to be accurate, the survey respondents must be representative of the population as a whole or the way in which the respondents differ from the population as a whole must be well understood so that estimates for the entire population can be made.

UNAIDS revised downward its estimate of the number of HIV-positive people in India from 5.7 million for 2006 to 2.5 million for 2007. This downward revision was not due to an actual enormous decline in HIV, but rather due to a change in the basis for the estimates. UNAIDS changed the basis of their estimates from clinic data for high-risk groups (pregnant women, injection drug users, commercial sex workers) to a more representative population-based survey. It was clear that the earlier estimates had greatly overestimated the prevalence of HIV in the general Indian population (Steinbrook, 2008; UNAIDS, 2007, 2008).

The new estimates are clearly more accurate than the old estimates. However, some are not happy about this change because they think it could lead to less attention and less money being allocated to fight HIV. Also, some interpret this reported change as real and thus exaggerate the extent of real declines in HIV. This example shows that how data are collected and how survey respondents are chosen for collection of the data can have a large impact on what conclusions are drawn.

The United Nations (1982) developed models of mortality patterns based on data from 22 less-developed countries. These new mortality models were intended to improve upon the earlier Coale–Demeny models that were mainly based on data from Europe or North America. They were intended to provide models of mortality that were more relevant to the situation in low-income countries. The United Nations developed a General Model based on data from all 22 countries, as well as four additional models based on data from subsets of the countries. Unfortunately, it was later concluded that the male model for the Latin American pattern was substantially a model of data error, due to problems with the data for males at the older ages in the contributing life tables (Dechter & Preston, 1991).

ASSESSMENT OF THE ACCURACY OF PAST PROJECTIONS, THE USE OF DATA TO DEVELOP MODELS FOR PROJECTING THE FUTURE, AND THE ASSUMPTIONS ON WHICH THOSE MODELS ARE BASED

Despite the importance of population projections, there has been fairly little work assessing their accuracy. Keilman's (1998) research is especially interesting. He assessed the accuracy of United Nations population projections for 1951–1988. Sometimes, the assessment was inaccurate because of error in the estimation of the population at the first time. After the results of the 1953 Chinese Census were released in 1954, the estimate of the population of the world for 1950 was increased, because it was seen that the population of China was more than 100 million larger than had earlier been thought. At other times, assumptions about the future were inaccurate. Throughout the world, mortality declined more rapidly than had been projected. Also, fertility declined more quickly after the 1970s than had been expected, partially owing to policies implemented because of alarm about high rates of population growth in many low-income countries in the 1960s and 1970s (Keilman, 1998).

Keilman (2008) also showed that in the period 1950–2001, population projection of the expected future demographic situation (total population, mortality, fertility, and international migration) done by European national statistical offices did not become more accurate. This was true even as the

amount and the quality of the data on which these forecasts were based improved.

The United Nations Population Division is the main producer of authoritative estimates and projections of the total population and of demographic processes, such as mortality and fertility. The most important part of a population projection is the future fertility assumption. Between 2004 and 2012, the UN Population Division changed its basic assumption about the course of fertility decline to a level that would result in zero population growth and about the fertility trajectory after that four times (Anderson, 2014; Basten, 2013b). With low mortality, zero population growth would require a total fertility rate (TFR) of 2.07. TFR resulting in eventual zero population growth is also called replacement fertility.<sup>1</sup>

These changes were based on observation of the history of some high-income, low-fertility countries. Before 2004, the UN Population Division had long projected TFR to asymptotically reach replacement level,  $TFR = 2.07$ . This assumed that all countries would eventually have low mortality and low fertility stationary populations.

In the 1990s, many countries had sustained below replacement fertility ( $TFR < 2.07$ ), sometimes falling to lowest-low fertility ( $TFR \leq 1.3$ ). After extensive consultation, fertility projection assumptions were changed in 2004, and all countries were then projected to asymptotically approach  $TFR = 1.86$ , which implies long-term population decline. This was a major departure from the earlier eventual zero population growth assumption.

Through the 2000s, TFR increased across at least two 5-year time periods in 21 below replacement fertility countries. On the basis of fertility increases in those countries, assumptions were again changed in 2010, so that in the new model TFR in below replacement fertility countries increased toward replacement, with the pace of increase more rapidly the farther TFR was below replacement. For countries with above replacement fertility in 2005–2010 such as Algeria, TFR was projected to fall below replacement and then increase toward replacement. This marked a return to the eventual stationary population assumption.

There was a loud outcry about the unreasonableness of the fertility projections for some Asian countries where fertility was very low and where there had been no indication of any increase (Basten, 2013a, 2013b; Basten, Coleman & Gu, 2012).

Partly in response to complaints about the 2010 estimates, in 2012, TFR projection assumptions were again changed. By 2012, TFR had increased across at least three 5-year periods in 25 low-fertility countries. The new low-fertility

---

1. The TFR is the number of children a woman would have in her life if she survived to age 50 and at every age had children at the rate in that population at that time.

projection model for many individual countries was based both on the experience of these 25 countries and on the TFR record of the individual country. The 2012 projections resulted in less extreme departures from earlier projected TFRs than occurred between the 2004 and 2010 projections.

### KEY ISSUES FOR FUTURE RESEARCH

Progress of the issues highlighted here requires scientific research but also discourse and discussion in philosophy and ethics. Just as in a trial, there is usually some merit on each side of an argument and there are conflicting considerations and values so that a perfect resolution in many areas is probably not possible.

#### AVAILABILITY AND ACCURACY OF DATA FOR NEW RESEARCH AND REANALYSIS WHILE PROTECTING HUMAN SUBJECTS

Despite rules by journals about access to data, Tenopir *et al.* (2011) document that data hoarding is still common. There needs to be consideration about what further steps could be taken while avoiding negative unintended consequences, such as discouraging data collection by researchers. Similarly, although there exists a procedure for external researchers to apply for use of INDEPTH data, the process is complicated and it is yet to be seen how open access will be.

The US Census Bureau has established Research Data Centers (RDCs) at more than 15 universities and research centers. At these centers, researchers with approved projects can research results of computer runs based on analysis of detailed individual data that are not available in as detailed a form in public use data sets. The RDCs help resolve the issue of data access and respondent confidentiality, but they are flawed by the requirement that projects using an RDC must “provide benefit to Census Bureau programs (U. S. Census Bureau, 2012).” This is an impediment to free scientific work and to the range of studies that can be pursued. More thought needs to go into this program, which seems to be affected by some of the same inclinations that have impeded data sharing among researchers.

The balance between data access and protection of respondents is a value-laden issue of public policy. More discussion between those concerned with research and those concerned with ethics could be fruitful to clarify what the guiding principles should be. These discussions seem even more necessary with increasing emphasis on “big data” to address many scientific and policy questions (Schuurman, 2000; United States. White House. Office of Science and Technology Policy, 2012).



PROBLEMS WITH THE ESTIMATION OF INDICATORS BASED ON NONGENERALIZABLE OR FLAWED DATA

The development of indicators and models, such as for mortality, that are relevant to the situation in parts of the world where data are lacking or of poor quality remains a challenge. There is an understandable urge to include all appropriate data in developing ways to make estimates, but there remains the danger of including data that include serious error.

The need for some basis for estimates, such as of mortality, is clear. A country wants to know things such as the average length of life (also called expectation of life at birth) for many purposes. However, in 2012, The United Nations Population Division reported that 26% of the countries in the world and 60% of the countries in Africa had no reliable data on adult mortality, and two countries had no reliable data on mortality at any age. In these situations, estimates based on the situation elsewhere are essential (United Nations, 2014, p. 14).

ASSESSMENT OF THE ACCURACY OF PAST PROJECTIONS, THE USE OF DATA TO DEVELOP MODELS FOR PROJECTING THE FUTURE, AND THE ASSUMPTIONS ON WHICH THOSE MODELS ARE BASED

How to use the past and thoughts about the future to model the future is a difficult problem. Additional assessment of the accuracy of past projections of the total population as well as of fertility and mortality could contribute to more informed decisions. In any case, probably, it is not wise to change assumptions frequently in a major way, as users of the results would easily assume real change where there has been none. In 2008, the UN projected that the TFR in Singapore in 2040–2045 would be 1.59; in 2010, with new fertility assumptions, it projected that the TFR in Singapore in 2040–2045 would be higher than was thought only 2 years earlier at 1.80; in 2012, the UN projected that the TFR in 2040–2045 would be lower than had been thought 2 years earlier at 1.38. Across the period 2008–2012, the estimated TFR in Singapore declined from 1.33 to 1.25. These changes in projected TFR had no relation to actual fertility changes in Singapore (Anderson, 2014). What changed over time was thinking about the future of fertility, based on the trajectory in less than  $\frac{1}{2}$  of the low-fertility countries rather than an observed change in many countries for which fertility was projected.

ACKNOWLEDGMENTS

Theresa Anderson, Heather King, Emily Marshall, Emily Merchant, John H. Romani, Michelle Steinmetz, and Victoria Velkoff have made helpful comments and suggestions.

## REFERENCES

- Alexander, J. T., Davern, M., & Stevenson, B. (2010). *Inaccurate age and sex data in the Census PUMS files: Evidence and implications*. National Bureau of Economic Research Working Paper No, 15703. Retrieved from <http://www.nber.org/papers/w15703.pdf>
- Anderson, B. A. (2014). *Projecting low fertility: Some thoughts about the plausibility and implications of assumptions*. University of Michigan Population Studies Center Research Report 14-815 (February). Retrieved from <http://www.psc.isr.umich.edu/pubs/pdf/rr14-815.pdf>
- Anderson, R. G., Greene, W. H., McCullough, B. D., & Vinod, H. D. (2008). The role of data/code archives in the future of economic research. *Journal of Economic Methodology*, 15, 99–119.
- Armstrong, M. P., Rushton, G., & Zimmerman, D. L. (1999). Geographically masking health data to preserve confidentiality. *Statistics in Medicine*, 18, 497–525.
- Baiden, F., Hodgson, A., & Binka, F. N. (2006). Demographic surveillance sites and emerging challenges in international health, Editorial. *Bulletin of the World Health Organization*, 86, 163–164.
- Basten, S. (2013a). Re-examining the fertility assumptions for Pacific Asia in the UN's 2010 World Population Prospects. University of Oxford Department of Social Policy and Intervention, *Barnett Papers in Social Research*: 2013/1 (June 7). Retrieved from 10.2139/ssrn.2275938
- Basten, S. (2013b). Comparing projection assumptions of fertility in six advanced Asian economies; or 'Thinking beyond the medium variant'. *Asian Population Studies*, 9, 322–331.
- Basten, S. A., Coleman, D. A., & Gu, B. (2012). Re-examining the fertility assumptions in the UN's 2010 World Population Prospects: Intentions and fertility recovery in East Asia. Paper presented at the Annual Meeting of the Population Association of America, San Francisco. Retrieved from <http://paa2012.princeton.edu/papers/122426>
- Berg, M. T., Stewart, E. A., Stewart, E., & Simons, R. L. (2013). A multilevel examination of neighborhood social processes and college enrollment. *Social Problems*, 60, 513–534.
- Caldwell, J., & Caldwell, P. (1988). Is the Asian family planning program model suited to Africa? *Studies in Family Planning*, 19, 19–28.
- Carrel, M., & Rennie, S. (2008). Demographic and health surveillance: Longitudinal ethical considerations. *Bulletin of the World Health Organization*, 86, 612–616.
- Chandramohan, D., Shibuya, K., Satel, P., Cairncross, S., Lopez, A. D., Murray, C. D. L., ... , Binka, F. (2008). Should data from demographic surveillance systems be made more widely available to researchers? *PLoS Medicine*, 5: 0169–0170.
- Coale, A. J., & Demeny, P. (1966). *Regional model life tables and stable populations*. Princeton, NJ: Princeton University Press.
- Dechter, A. R., & Preston, S. H. (1991). Age misreporting and its effect on adult mortality estimates in Latin America. *Population Bulletin of the United Nations*, 31(32), 1–16.

- Fairchild, A. L., & Bayer, R. (1999). Uses and abuses of Tuskegee. *Science*, *284*, 919–921.
- Fisher, T. L. (2010). The income of the elderly: The effect of changes to reported age in the Current Population Survey. Paper presented at the annual conference of the Association for Public Policy Analysis and Management, Boston, October 13. Retrieved from <https://www.appam.org/conferences/fall/boston2010/sessions/downloads/4555.1.pdf>
- Freedman, D., Thornton, A., Camburn, D., Alwin, D., & Young-DeMarco, L. (1988). The life history calendar: A technique for collecting retrospective event-history data. *Sociological Methodology*, *18*, 37–68.
- INDEPTH Network (2014). *INDEPTH Network: Better health data for better health policy*. Retrieved from <http://www.indepth-network.org/>
- Jones, J. H. (1981). *Bad blood: The Tuskegee syphilis experiment*. New York, NY: The Free Press.
- Keilman, N. (1998). How accurate are the United Nations world population projections? *Population and Development Review*, *24* (Supplement: Frontiers of Population Forecasting), 15–41.
- Keilman, N. (2008). European demographic forecasts have not become more accurate over the past 25 years. *Population and Development Review*, *34*, 137–153.
- McCullough, B. D. (2009). Open access economics journals and the market for reproducible economic research. *Economic Analysis & Policy*, *39*, 117–126.
- McCullough, B. D., McGeary, K. A., & Harrison, T. D. (2008). Do economics journal archives promote replicable research? *Canadian Journal of Economics*, *41*, 1406–1420.
- Sherman, J. E., & Fetters, T. L. (2007). Confidentiality concerns with mapping survey data in reproductive health research. *Studies in Family Planning*, *38*, 309–321.
- Schuurman, N. (2000). Trouble in the heartland: GIS and its critics in the 1990s. *Progress in Human Geography*, *24*, 569–590.
- Steinbrook, R. (2008). HIV in India—A downsized epidemic. *New England Journal of Medicine*, *358*, 107–109.
- Tenopir, C., Allard, S., Douglas, K., Aydinoglu, A. U., Read, E., Manoff, M., & Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PloS One*, *6*, 1–21.
- United Nations (1982). *Model life tables for developing countries*. New York, NY: United Nations.
- United Nations (2014). *World population prospects: The 2012 revision, Methodology of the United Nations population estimates and projections*, New York: United Nations. Retrieved from [http://esa.un.org/unpd/wpp/Documentation/pdf/WPP2012\\_Methodology.pdf](http://esa.un.org/unpd/wpp/Documentation/pdf/WPP2012_Methodology.pdf)
- UNAIDS (2007). *2.5 million people living with HIV in India*. Retrieved from [http://www.unaids.org/en/KnowledgeCentre/Resources/FeatureStories/archive/2007/20070704\\_India\\_new\\_data.asp](http://www.unaids.org/en/KnowledgeCentre/Resources/FeatureStories/archive/2007/20070704_India_new_data.asp)
- UNAIDS website (2008). *Q + A on India's revised AIDS estimates*. Retrieved from [http://data.unaids.org/pub/InformationNote/2007/070701\\_india%20external\\_qa\\_en.pdf](http://data.unaids.org/pub/InformationNote/2007/070701_india%20external_qa_en.pdf)

- U. S. Census Bureau (2010). *Analysis of perturbed and unperturbed age estimates: 2008*. Retrieved from [http://www.census.gov/cps/user\\_note\\_age\\_estimates.html](http://www.census.gov/cps/user_note_age_estimates.html)
- U. S. Census Bureau (2012). *RDC research opportunities*. Center for Economic Studies. Retrieved from <https://www.census.gov/ces/rdcresearch/>
- United States. White House. Office of Science and Technology Policy (2012). *Obama administration unveils "Big Data" initiative: Announces \$200 million in new R&D investments*. Retrieved at [http://www.whitehouse.gov/sites/default/files/microsites/ostp/big\\_data\\_press\\_release\\_final\\_2.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf)

### FURTHER READING

- Coale, A. J., & Trussell, T. J. (1996). The development and use of demographic models. *Population Studies*, 50, 469–484.
- National Academy of Science (1995). *On being a scientist: Responsible conduct of research* (2nd ed.). Washington, DC: National Academies Press.
- Preston, S. H. (1993). The contours of demography: Estimates and projections. *Demography*, 30, 593–606.
- United States, National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (1979). *The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research*. Retrieved from [http://www.fda.gov/ohrms/dockets/ac/05/briefing/2005-4178b\\_09\\_02\\_Belmont%20Report.pdf](http://www.fda.gov/ohrms/dockets/ac/05/briefing/2005-4178b_09_02_Belmont%20Report.pdf).
- VanWey, L., Rindfuss, R. R., Gutman, M. P., Entwisle, B., & Balk, D. L. (2005). Confidentiality and spatially explicit data: Concerns and challenges. *Proceedings of the National Academy of Sciences of the United States of America*, 102: 15337–15342. Retrieved from <http://www.pnas.org/content/102/43/15337.full>

### BARBARA A. ANDERSON SHORT BIOGRAPHY

**Barbara A. Anderson** is the Ronald Freedman Collegiate Professor of Sociology and Population Studies at the University of Michigan. She holds an AB degree in mathematics from the University of Chicago and a PhD degree in sociology from Princeton University. She has been a faculty member at Yale University and Brown University, a visiting member at the Institute for Advanced Study, and a fellow at the Center for Advanced Study in the Behavioral Sciences. She has conducted extensive research on the relation of population and development and the role of data and data quality in these areas. She has consulted on data and research with the governments of Estonia, China, and South Africa. She is a member of the U.S. Census Scientific Advisory Committee and has served on the NSF Review Panel on Sociology and the NICHD Population Research Committee. She has published or edited six books and more than 100 articles and chapters.

RELATED ESSAYS

Expertise (*Sociology*), Gil Eyal

The Evidence-Based Practice Movement (*Sociology*), Edward W. Gondolf

Globalization of Capital and National Policymaking (*Political Science*),  
Steven R. Hall

Trends in the Analysis of Interstate Rivalries (*Political Science*), William R.  
Thompson

Political Inequality (*Sociology*), Jeff Manza

Why Do Governments Abuse Human Rights? (*Political Science*), Will H.  
Moore and Ryan M. Welch

Causation, Theory, and Policy in the Social Sciences (*Sociology*), Mark C.  
Stafford and Daniel P. Mears

Trends in the Analysis of Interstate Rivalries (*Political Science*), William R.  
Thompson

Translational Sociology (*Sociology*), Elaine Wethington