# Demography: Past, Present, and Future

Yu Xie

In a classic statement, Hauser and Duncan (1959) defined demography as "the study of the size, territorial distribution, and composition of population, changes therein, and the components of such changes" (p. 2). It was fortunate that Hauser and Duncan explicitly included "composition of population" and "changes therein" in their definition, for their inclusion has broadened demography to encompass two types of demography: formal demography and population studies. Formal demography, whose origin can be traced to John Graunt in 1662, is concerned with fertility, mortality, age structure, and spatial distribution of human populations. Population studies is concerned with population compositions and changes from substantive viewpoints anchored in another discipline, be it sociological, economic, biological, or anthropological; its origin can be traced to Thomas Malthus in 1798. By definition, population studies is interdisciplinary, bordering between formal demography and a substantive discipline that is often, but not necessarily, a social science.

Defined in this way, demography provides the empirical foundation on which other social sciences are built. It is hard to imagine that a social science can advance steadily without first knowing the basic information about the human population that it studies. As a field of inquiry, demography enjoyed a rapid growth in the twentieth century. For example, the membership of the Population Association of America (PAA), the primary association for demographers in the U.S. founded in 1931, grew from fewer than 500 in 1956 to more than 3,000 in 1999. This growth is remarkable given the virtual absence of demography departments at American universities (with a few exceptions, such as the University of California Berkeley). To recognize contributions made by demographers, one only needs to be reminded of factual information about contemporary societies. Much of what we know as "statistical facts" about American society, for instance, has been provided or studied by demographers. Examples include socioeconomic inequalities by race (Farley 1984) and gender (Bianchi and Spain 1986), residential segregation by race (Duncan 1957; Massey and Denton 1993), intergenerational social mobility (Blau and Duncan 1967; Featherman and Hauser 1978), increasing trends of divorce (Sweet and Bumpass 1987) and cohabitation (Bumpass, Sweet, and Cherlin 1991), consequences of single parenthood for children (McLanahan and Sanderfur 1994), rising income inequality (Danziger and Gottschalk 1995), and increasing economic returns for college education (Mare 1995).

Besides providing factual information, demography has also been fundamental in forecasting future states of human societies. Although demographic forecasting is subject to uncertainty, as any type of forecasting, demographers are able to predict future population sizes by age with a high degree of confidence, utilizing information pertaining to past fertility, regularity in age patterns of mortality, and likely future levels of mortality. A notable example of demographic forecasting is the work by Lee and Tuljapurkar (1994, 1997), who demonstrated how demographic forces (i.e., projected improvements in longevity) dramatically impact future demands on social security.

Formal demography and population studies not only take on different subject matters, but also rely on different methodological approaches. Characteristically, formal demography is built on mathematics and thus is closely tied to mathematical demography. It has a rich arsenal of powerful research tools, such as life tables and stable population theory, the latter of which is usually accredited to Alfred J. Lokta in 1922. Note that mathematical models in formal demography sometimes incorporate stochastic processes. The refinement and formalization of mathematical demography and its successful application to human populations can be found in works by Coale (1972), Keyfitz (1985), Preston and Campbell (1993), Rogers (1975), and Sheps and Menken (1973). In its applications, mathematical demography typically presumes access to population data and handles heterogeneity through disaggregation (i.e., dividing a population into subpopulations).

Methods used in population studies are eclectic, borrowing heavily from substantive social science disciplines. Given the widespread use of survey data and the predominant role of statistical inference in all social science disciplines since the 1960s, it should come as no surprise that the characteristic method in population studies is statistical. [It should be noted that qualitative methods are also used in demography (Kertzer and Fricke 1997).] The types of statistical methods used by demographers vary a great deal and change quickly, ranging from path analysis and structural equations (Duncan 1975) and log-linear models (Goodman 1984) to econometric models (Heckman 1979; Willis and Rosen 1979) and event history models (Yamaguchi 1991). Substantive research in population studies usually involves statistical analysis of sample data (as opposed to population data) in a multivariate framework; sometimes, researchers develop statistical models to test hypotheses derived from an individual-level behavioral model. For this reason, population studies is closely tied to statistical demography.

Although it is useful to draw the distinction between formal demography and population studies, the boundary between the two is neither fixed nor impermeable. Indeed, this

Yu Xie is Frederick G. L. Huetwell Professor of Sociology, Department of Sociology, University of Michigan, Ann Arbor, MI 48109 (E-mail: *yuxie@umich.edu*). The author thanks Jennifer Barber, Mark Becker, Aimee Dechter, Tom Fricke, Robert Hauser, David Harris, Robert Mare, and Pam Smock for their comments on an earlier draft of this article.


Journal of the American Statistical Association
June 2000, Vol. 95, No. 450, Vignettes

boundary presents many exciting topics for research. For example, it is possible, and indeed desirable, to estimate demographic rates from sample data with statistical models before feeding them as input for mathematical analysis in formal demography (e.g., Clogg and Eliason 1988; Hoem 1987; Land, Guralnik, and Blazer 1994; Xie 1990; Xie and Pimentel 1992). There are good reasons for using statistical tools in combination with mathematical models. First, the advancement of demography has brought with it more, richer, and better data in the form of sample data; the use of sample data requires statistical tools; and treating sample data as exactly known quantities runs the danger of being contaminated by sampling errors. Second, statistical models are better suited for examining group differences through the use of covariates, for the method of disaggregation presumes a full interaction model and may lead to inaccurate estimation due to small group sizes. Third, because observed data may sometimes be irregular or simply missing, statistical models can help smooth or impute data (e.g., Little and Rubin 1987). Conversely, techniques in formal demography (e.g., indirect estimation, intercensal cohort comparisons, model life tables) can also be utilized to improve statistical analysis. For example, mathematical relationships for demographic components are often used to correct faulty data or provide missing data before they are analyzed with statistical techniques.

Let me now turn to the methodological implications of Hauser and Duncan's two important phrases "composition of population" and "changes therein" in their definition of demography. The first phrase refers to population heterogeneity; the second, to dynamic processes. Population heterogeneity or variability is fundamental to all social science. As articulated by Mayr (1982), one of the most important and long-lasting influences of Darwin's evolution theory is "population thinking," as distinguished from "typological thinking" prevalent in physical science. Briefly stated, typological thinking attempts to discover the "essence" or "truth" in society and human nature. Once well understood, a scientific concept or law is always generalizable to all settings at all times. If the observed phenomena appear in disarray, then averages are taken and deviations are disregarded. In the history of statistics, typological thinking is associated with the "average man" in Quetelet's social physics. In contrast, population thinking makes *variations* the very subject matter to be studied. It was Francis Galton, Darwin's cousin, who introduced population thinking into the field of statistics and in so doing discovered correlation and regression (see Xie 1988 and references therein). Demography should pride itself for approaching social phenomena through population thinking by focusing on variations by group and individual characteristics. For example, demographers have long been interested in how life chances differ by gender, age, race, region, and family origin, as well as historical and cultural context.

In studying social changes, demographers have advocated the perspectives of cohort (Ryder 1965) and life course (Elder 1985). Whereas the latter is concerned with the occurrences of significant events and transitions over individuals' lives and the dependence of these events and transitions on earlier experiences and societal forces, the former relates individuals' different experiences at the micro level to social changes at the macro level. That is, demographers are especially interested in both age-graded intracohort changes and intercohort changes resulting from cohort replacement. In addition, demographers also accept the notion that certain temporal effects are period based, affecting all individuals regardless of age. Due to the linear dependency of the three measures, however, the conceptual model of age, period, and cohort is intrinsically underidentified, and its statistical implementation requires constraining assumptions based on substantive considerations (Mason and Fienberg 1985).

Many of the advances in demography over the past few decades cannot be separated from the field's close relationship to statistics. Indeed, it is the integration of statistical methods into demographic research that has enabled demographers to study population heterogeneity and change with sample data. A prime example where statistical methodology has significantly impacted demography is event history analysis (also called survival models or hazard models). As argued by Tuma and Hannan (1984), event history analysis is ideally suited for studying social dynamics and social processes, because it deals with the timing of transitions or event occurrences. Although event history analysis was invented by demographers in the form of life tables, the life table approach in mathematical demography assumes population data and cannot easily incorporate population heterogeneity except by disaggregation. Cox's (1972) influential work bridged the gap between life tables and regression analysis by focusing on estimation of the effects of covariates using sample data. The influence of Cox (1972) in demography goes beyond his innovative estimation method, for it represents a shift in methodological orientation toward statistical demography. With this new orientation, event history analysis is seen as a statistical operationalization of life tables for sample data, with attention paid primarily to group differences or the effects of covariates.

Studying population heterogeneity by observed group or individual characteristics with event history analysis has proven fruitful; see, for example, Thornton, Axinn, and Teachman's (1995) study of the effects of education on the likelihood of cohabitation and marriage and Wu's (1996) work on the determinants of premarital childbearing. In addition, demographers have a long-standing concern with unobserved heterogeneity in event history analysis and life tables (e.g., Sheps and Menken 1973). Unlike in the case of linear models, unobserved heterogeneity uncorrelated with covariates in event history data could bias estimated hazards through altering the composition of the exposed population at risk (Vaupel and Yashin 1985). For example, demographers have long been puzzled as to whether the well-documented black/white mortality crossover is due to biological selection, in which old blacks are more fit than old whites (Manton, Poss, and Wing 1979), or due to age misreporting (Preston, Elo, Rosenwaike, and Hill 1996). There are two broad approaches to handling unobserved heterogeneity: parametric mixture models (see Powers and

Xie 2000) and Heckman and Singer's (1984) nonparametric method that is analogous to latent class models for contingency tables. To improve identification, demographers have also capitalized on pooled information from clustered data structures, such as siblings or twins, assuming unobserved heterogeneity at the cluster level instead of at the individual level (Guo and Rodriguez 1992; Yashin and Iachine 1997).

As stated earlier, population studies is interdisciplinary, with sociological and economic demography taking center stage. In both sociology and economics, there has been a methodological tradition of structural equation models simultaneously representing multiple processes with strong theoretical priors. Event history analysis has been incorporated into this framework, with the hazard rate treated as a limited dependent variable. In this framework, demographers have gone beyond assuming unobserved heterogeneity orthogonal to covariates. Instead, they consider unobserved heterogeneity that is selective with respect to observed variables (Lillard, Brien, and Waite 1995). But identification of such models is difficult and requires pooled information from repeated events and strong parametric assumptions.

Where will demography go from here? Although predicting the future inherently carries some risks, past experience provides some basis for making cautionary conjectures about demography's future. First, demography, especially in the United States, has been a very successful interdisciplinary enterprise and will continue to be one. Not only will it retain sociologists and economists who already strongly identify themselves with demography, but it will also attract the interest of scholars in other social and biological sciences. In particular, incorporation of biological approaches in demographic research looks promising (for a recent example, see Yashin and Iachine 1997). Second, the interdisciplinary nature of demography makes it an ideal locus for innovating, testing, and popularizing new statistical methods. This means closer ties between statistical science and demography, as statistical demography continues to affirm its prominent role in demography. Third, in the burgeoning "information age," the public will demand more easily accessible information about the society in which they live, as exemplified by the strong public interest in the implementation of the 2000 U.S. Census. Demography can play an important role in providing that information. The internet is and will continue to be a powerful tool for bringing demographic information to the general public and policymakers alike.

One substantive area where demography is rapidly making progress is in research on aging. As the elderly comprise an increasingly large portion of the population in many societies, including the United States, demographers have become increasingly interested in aging (e.g., Wachter and Finch 1997). Economists and sociologists are already busy studying social and economic aspects aging, such as retirement, health, residence, economic security, and family support. The demography of aging has also benefited from contributions from scholars in biology, psychology, public health, gerontology, and geriatrics. This is an area where the aforementioned conjectures about the future of demography will likely be manifested, as we will see in the demography of aging more interdisciplinary work, newer and better applications of statistical methodology, and provision of more public and scientific information.

## REFERENCES


Bianchi, S. M., and Spain, D. (1986), *American Women in Transition*, New York: Russell Sage Foundation.

Blau, P. M., and Duncan, O. D. (1967), *The American Occupational Structure*, New York: Wiley.

Bumpass, L. L., Sweet, J. A., and Cherlin, A. (19 ), "The Role of Cohabitation in Declining Rates of Marriage," *Journal of Marriage and the Family*, 53, 913–927.

Clogg, C. C., and Eliason, S. R. (1988), "A Flexible Procedure for Adjusting Rates and Proportions, Including Statistical Methods for Group Comparisons," *American Sociological Review*, 53, 267–283.

Coale, A. J. (1972), *The Growth and Structure of Human Populations: A Mathematical Investigation*. Princeton, NJ: Princeton University Press.

Cox, D. R. (1972), "Regression Models and Life-Tables," *Journal of the Royal Statistical Society*, Ser. B, 34, 187–220.

Danziger, S., and Gottschalk, P. (1995), *America Unequal*, New York: Russell Sage.

Duncan, O. D. (1957), *The Negro Population of Chicago: A Study of Residential Succession*, Chicago: University of Chicago Press.

——— (1975), *Introduction to Structural Equation Models*, New York: Academic Press.

Elder, G. H., Jr. (1985), "Perspectives on the Life Course," in *Life Course Dynamics: Trajectories and Transitions, 1968–1980*, ed. G. H. Elder, Jr., Ithaca, NY: Cornell University Press, pp. 23–49.

Farley, R. (1984), *Blacks and Whites: Narrowing the Gap?*, Cambridge, MA: Harvard University Press.

Featherman, D. L., and Hauser, R. M. (1978), *Opportunity and Change*, New York: Academic Press.

Goodman, L. A. (1984), *The Analysis of Cross-Classified Data Having Ordered Categories*, Cambridge, MA: Harvard University Press.

Guo, G., and Rodriguez, G. (1992), "Estimating a Multivariate Proportional Hazards Model for Clustered Data Using the EM Algorithm, With an Application to Child Survival in Guatemala," *Journal of the American Statistical Association*, 87, 969–976.

Hauser, P. M., and Duncan, O. D. (Eds.) (1959), *The Study of Population: An Inventory and Appraisal*, Chicago: University of Chicago Press.

Heckman, J. J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153–161.

Heckman, J. J., and Singer, B. (1984), "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52, 271–320.

Hoem, J. M. (1987), "Statistical Analysis of a Multiplicative Model and Its Application to the Standardization of Vital Rates: A Review," *International Statistical Review*, 55, 119–152.

Keyfitz, N. (1985), *Applied Mathematical Demography* (2nd ed.), New York: Springer-Verlag.

Kertzer, D. I., and Fricke, T. (Eds.) (1997), *Anthropological Demography: Toward a New Synthesis*, Chicago: University of Chicago Press.

Land, K. C., Guralnik, J. M., and Blazer, D. G. (1994), "Estimating Increment-Decrement Life Tables With Multiple Covariates From Panel Data: The Case of Active Life Expectancy," *Demography*, 31, 297–319.

Lee, R. D., and Tuljapurkar, S. (1994), "Stochastic Population Forecasts for the United States: Beyond High, Medium, and Low," *Journal of the American Statistical Association*, 89, 1175–1189.

——— (1997), "Death and Taxes: Longer Life, Consumption, and Social Security," *Demography*, 34, 67–81.

Lillard, L. A., Brien, M. J., and Waite, L. J. (1995), "Premarital Cohabitation and Subsequent Marital Dissolution: A Matter of Self-Selection?," *Demography*, 32, 437–457.

Little, R. J. A., and Rubin, D. B. (1987), *Statistical Analysis With Missing Data*, New York: Wiley.

Manton, K. G., Poss, S. S., and Wing, S. (1979), "The Black/White Mortality Crossover: Investigation From the Perspective of the Components of Aging," *The Gerontologist*, 19, 291–300.

Mare, R. D. (1995), "Changes in Educational Attainment and School Enrollment," in *State of the Union: America in the 1990s*, Vol. II, ed. R. Farley, New York: Russell Sage.

Mason, W. M., and Fienberg, S. E. (1985), *Cohort Analysis in Social Research: Beyond the Identification Problem*, New York: Springer-Verlag.

Massey, D. S., and Denton, N. A. (1993), *American Apartheid: Segregation and the Making of the Underclass*, Cambridge, MA: Harvard University Press.

Mayr, E. (1982), *The Growth of Biological Thought: Diversity, Evolution, and Inheritance*, Cambridge, MA: Belknap Press.

McLanahan, S. S., and Sandefur, G. (1994), *Growing Up With a Single Parent: What Hurts, What Helps*, Cambridge, MA: Harvard University Press.

Powers, D. A., and Xie, Y. (2000), *Statistical Methods for Categorical Data Analysis*, New York: Academic Press.

Preston, S. H., and Campbell, C. (1993), "Differential Fertility and the Distribution of Traits: The Case of IQ," *American Journal of Sociology*, 98, 997–1019.

Preston, S. H., Elo, I. T., Rosenwaike, I., and Hill, M. (1996), "African-American Mortality at Older Ages: Results of a Matching Study," *Demography*, 33, 193–209.

Rogers, A. (1975), *Introduction to Multiregional Mathematical Demography*, New York: Wiley.

Ryder, N. B. (1965), "The Cohort as a Concept in the Study of Social Change," *American Sociological Review*, 30, 843–861.

Sheps, M. C., and Menken, J. A. (1973), *Mathematical Models of Conception and Birth*, Chicago: University of Chicago Press.

Sweet, J. A., and Bumpass, L. L. (1987), *American Families and Households*, New York: Russell Sage.

Thornton, A., Axinn, W., and Teachman, J. D. (1995), "The Influence of School Environment and Accumulation on Cohabitation and Marriage in Early Adulthood," *American Sociological Review*, 60, 762–774.

Tuma, N. B., and Hannan, M. T. (1984), *Social Dynamics: Models and Methods*, San Francisco: Academic Press.

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.

Wachter, K. W., and Finch, C. E. (Eds.) (19 ), *Between Zeus and the Salmon*, Washington, DC: National Academy Press.

Willis, R. J., and Rosen, S. (1979), "Education and Self-Selection," *Journal of Political Economy*, 87(5), S7–S36.

Wu, L. (1996), "Effects of Family Instability, Income, and Income Instability on the Risk of a Premarital Birth," *American Sociological Review*, 61, 386–406.

Xie, Y. (1988), "Franz Boas and Statistics," *Annals of Scholarship*, 5, 269–296.

——— (1990), "What is Natural Fertility? The Remodelling of a Concept," *Population Index*, 56, 656–663.

Xie, Y., and Pimentel, E. E. (1992), "Age Patterns of Marital Fertility: Revising the Coale–Trussell Method," *Journal of the American Statistical Association*, 87, 977–984.

Yamaguchi, K. (1991), *Event History Analysis*, London: Sage.

Yashin, A. I., and Iachine, I. I. (1997), "How Frailty Models can be Used for Evaluating Longevity Limits: Taking Advantage of an Interdisciplinary Approach," *Demography*, 34, 31–48.