Small-Area Analysis

Stanley K. Smith

"Small-Area Analysis," pages 898-901 in Paul Demeny and Geoffrey McNicoll (eds.), *Encyclopedia of Population.* Farmington Hills, MI: Macmillan Reference, 2003.

Small-Area Analysis

Stanley K. Smith

Spurred by new business applications and government programs, the demand for small-area demographic data and analysis has grown tremendously in recent decades. To meet that demand, analysts have drawn on an expanding set of data sources, statistical techniques, and computer applications. The result has been improved data quality across a broad spectrum of variables and geographic areas, enhancing both the usefulness and the importance of small-area analyses.

There are no hard-and-fast rules for defining *small areas*. Under one definition, they are states and other subnational areas for which area-specific samples from national surveys are too small to provide meaningful estimates (Rao, 1999). More typically, *small areas* refer to counties and subcounty areas like cities, census tracts, ZIP code areas, and even individual blocks (Murdock and Ellis, 1991). They range from less than an acre to thousands of square miles, and from no inhabitants to many millions.

This article reviews commonly used data sources and application techniques and discusses several distinct features of small-area demographic analysis. Although reference is primarily to the United States, many of the issues discussed transcend national boundaries. Analysts using small-area data include demographers, sociologists, geographers, economists, marketers, epidemiologists, planners, and others.

Data Sources

***Censuses*.** Most industrialized countries conduct reliable censuses of the entire population at regular intervals (e.g., every five or 10 years). In developing countries, censuses are neither as frequent nor as reliable. In most countries, however, censuses constitute the most

comprehensive source of small-area data, typically covering a variety of population characteristics (e.g., age, sex, race, marital status, income, education) and housing information (e.g., number of units in structure, year built, number of rooms, value or monthly rent).

Administrative Records. Censuses generally provide accurate and comprehensive data, but only infrequently. Administrative records kept by federal, state, and local governments provide small-area data for years after or between censuses. These records provide information on variables such as births, deaths, school enrollment, social insurance, building permits, drivers licenses, voter registration, and income tax returns—each reflecting a facet of population structure and change that may be useful for constructing estimates and projections or for tracking demographic trends. Most industrialized countries maintain relatively accurate records; some European countries even produce census statistics based *solely* on administrative records (Longva, Thomsen, and Severeide, 1998). In many developing countries, however, administrative records are plagued by incomplete coverage (Cleland, 1996).

Sample Surveys. Sample surveys are another common source of demographic and socioeconomic data. Their principal limitation is that sample sizes are generally too small to provide reliable estimates for small areas. A notable exception is the American Community Survey, which is expected to cover some three million U.S. households annually by 2003 and eventually to generate useful estimates (based on three to five years of data) down to the block group level for the entire nation.

Estimates and Projections

Censuses, administrative records, and sample surveys are the major sources of data for small-area demographic analyses. For many purposes, however, these data must be transformed into population estimates or projections. Population *estimates* refer to current or past time

periods for which census data are unavailable and typically derive from methods that integrate census, administrative record, and/or sample survey data. Population *projections* refer to future time horizons and are based on historical trends and assumptions regarding future trends.

***Estimates*.** Small-area estimates of total population generally rely on housing unit, component, or regression methods (Murdock and Ellis, 1991). The *Housing Unit* method derives population estimates from calculations of occupied housing units (i.e., households) and average household size, plus the number of persons living in group quarters facilities (e.g., college dormitories, military barracks, prisons). *Component methods* derive population estimates from birth, death, and migration data (births and deaths from vital statistics records, migration from changes in school enrollments, income tax returns, or other symptomatic variables). *Regression methods* derive population estimates from symptomatic indicators of population change (e.g., births, school enrollment, electric utility customers, registered voters, drivers licenses, tax returns) in a multivariate model. All three types of methods produce useful estimates, but the housing unit method is the most commonly used for small-area estimates because the requisite data are so widely available.

Estimates of demographic characteristics such as age, sex, and race are typically based on the cohort-component method (Siegel, 2001). Here, birth, death, and migration rates are applied separately to each age/sex/race subgroup in the population. Estimates of socioeconomic characteristics such as income, employment, and education are often based on synthetic techniques, whereby known proportions of the population exhibiting a characteristic in a larger area (e.g., a state) are applied to population estimates for smaller areas (e.g., cities, counties); typically, these proportions are calculated separately for different subgroups of the population

(e.g., age, sex, race). Estimates of demographic and socioeconomic characteristics can also be based on administrative records (e.g., Fonseca and Tayman, 1989).

Projections. Most population projection methods can be grouped into three basic categories (Smith, Tayman, and Swanson, 2001). *Trend extrapolation* methods extend observed historical trends. These methods may be simple (e.g., past growth rates projected to remain constant) or complex (e.g., ARIMA time series models). Trend extrapolation methods are frequently used for small-area projections because their data requirements are minimal, they are easy to apply, and their forecasts have often proven to be reasonably accurate.

The *cohort-component* method accounts separately for the three components of population change (births, deaths, and migration). Applications of this method divide the population into age-sex subgroups (i.e., cohorts), sometimes further distinguished by race, ethnicity, or other demographic characteristics. Projections of each component can be based on the extrapolation of past trends, projected trends in other areas, structural models, or professional judgment. Simplified versions of the method can also be applied (e.g., Hamilton and Perry, 1962). The cohort-component method is the most frequently used projection method because it incorporates a broad range of data sources, assumptions, and application techniques and can provide projections of demographic characteristics as well as total population.

Structural models are based on an entirely different logic. They relate population projections to variables known to drive population change (e.g., comparative wages, employment, land use). Some are simple, involving only a single equation and a few variables; others are much more complex, containing many equations, variables, and parameters. Structural models typically differentiate among individual components of growth and are often used in combination with the cohort-component method. Although they require highly detailed

data and substantial investments of effort and modeling skill, structural models provide a broader range of projections than other methods.

Uses of Small-Area Data

***Increase Knowledge*.** Small-area data have several uses. One is to raise understanding of socioeconomic and demographic variations across geographic areas (e.g., states, counties, cities, census tracts). Empirical studies have investigated how differences in government tax and expenditure policies affect interprovincial migration in Canada (Day, 1992); how poverty, urbanization, and geographic location affect the incidence of cholera among regions in Mexico (Borroto and Martinez-Piedra, 2000); and how differences in industrial, educational, geographic, and demographic characteristics affect county population growth rates in the United States (Beeson, DeJong, and Troesken, 2001). The primary purpose of these studies is to enlarge the body of knowledge available to scholars, policy makers, and other analysts.

***Inform Public Policy*.** Another use of small-area data is to support decision making by federal, state, and local government agencies. Small-area data are indispensable for drawing political and electoral boundaries, allocating government funds, siting public facilities, developing program budgets, determining eligibility for public programs, and monitoring program effectiveness. Examples include using block-level population and household projections to choose sites for fire stations (Tayman, Parrott, and Carnevale, 1997); using birth data by ZIP code area to identify areas in need of adolescent pregnancy prevention programs (Gould, Herrchen, Pham, Bera, and Brindis, 1998); and developing techniques for monitoring infectious diseases by health district (Hashimoto, Murakami, Taniguchi, and Nagai, 2000).

Support Business Decision Making. Small-area data figure prominently in many types of business decisions (e.g., site selection, sales forecasting, consumer profiles, litigation support,

target marketing, and labor force analysis). Analyses of this type parallel those that inform public policy, but their focus is on private rather than public sector decision making. Examples include using small-area demographic data to select locations for supermarkets (Morrison and Abrahamse, 1996); projecting the number of births to determine a hospital's need for additional obstetrical services (Thomas, 1997); and using population estimates and projections to support a company's bank loan application (Murdock and Hamm, 1997).

Distinctive Problems of Small-Area Analysis

Several distinctive problems of small-area analysis require special attention. First, unlike states and most counties, the geographic boundaries of many subcounty areas do not remain constant over time. Cities annex adjoining areas, census tracts get subdivided, ZIP code areas are reconfigured, service areas are redefined, and new statistical areas are established. All such changes undermine the consistency of historical data series.

Second, many types of data are not tabulated for areas below a certain level of geography. Consequently, analyses routinely performed for larger areas (e.g., states) may be impossible (or feasible only using proxy variables) for smaller areas (e.g., census tracts).

Third, errors plague even the best censuses, estimates, projections, and administrative record systems. Their impact on data reliability is typically greater for small areas than large areas, where errors are often mutually offsetting. In addition, survey data are generally less reliable for small areas than large areas because sample sizes are smaller and survey responses more variable.

Finally, particular events are more likely to disrupt orderly trends for small areas than large areas. The opening or closing of a prison or military base, the construction of a large housing development, the opening of a new road or railway, and the addition or loss of a major

employer may sharply distort trends in a specific area within a city, county, or state. The impact of growth constraints (e.g., swamps, zoning restrictions) and seasonal populations (e.g., snowbirds, migrant workers) is also greater for small areas than large areas. Whereas the effects of these idiosyncratic factors tend to cancel each other out and generally can be ignored in empirical analyses for large areas, such is not the case for small areas.

Conclusions

Small-area analyses support both public and private sector decision making. They enhance the development and evaluation of public policies and help target resources more effectively to areas in need. They help businesses select store sites, devise strategic plans, profile potential consumers, target local markets, and forecast future earnings. Small-area data provide the raw materials for constructing a variety of widely used measures for uniquely defined geographic areas. Despite their limitations, small-area data and analysis are essential for a wide variety of purposes.

Several recent developments have broadened the scope and advanced the quality of small-area analyses. A wider variety of small-area data have become available in many countries, primarily through administrative records and sample surveys. The Internet has greatly enhanced access to these data, and an explosion of computing power, data storage capacity, and software applications has expanded their potential usefulness. Geographic information systems (GIS) technology has facilitated the collection, organization, manipulation, analysis, and presentation of geographically-referenced data. These developments have prompted many new uses of small-area data, at ever lower levels of geography.

Concerns about privacy and confidentiality, however, pose a formidable barrier to the continued advancement of small-area

analysis. To many, the collection of personal information is an invasion of privacy. Confidentiality is potentially at risk when personal data are shared among public and private agencies. Such concerns have caused many statistical agencies to curtail the release of demographic data and have restricted the use of administrative records in the United States, Germany, the United Kingdom, and elsewhere (Doyle, Lane, Theeuwes, and Zayatz, 2001; Redfern, 1989). Devising acceptable ways to utilize information, while preserving privacy and confidentiality, is a major challenge for small-area analysts.

See also BUSINESS DEMOGRAPHY; POPULATION PROJECTIONS; STATE AND LOCAL DEMOGRAPHY.

ACKNOWLEDGEMENTS: The author thanks Peter Morrison, Linda Jacobsen, and Jeff Tayman for numerous helpful suggestions.

REFERENCES

Beeson, Patricia E., David N. DeJong, and Werner Troesken. 2001. "Population Growth in U.S. Counties, 1840-1990." *Regional Science and Urban Economics* 31: 669-699.

Borroto, Rene J. and Ramon Martinez-Piedra. 2000. "Geographical Patterns of Cholera in Mexico, 1991-1996." *International Journal of Epidemiology* 29: 764-772.

Cleland, John. 1996. "Demographic Data Collection in Less Developed Countries 1946-1996." *Population Studies* 50: 433-450.

Day, Kathleen M. 1992. "Interprovincial Migration and Local Public Goods." *Canadian Journal of Economics* 25: 123-144.

Doyle, Pat, Julia I. Lane, Jules J.M Theeuwes, and Laura V. Zayatz (eds.). 2001. *Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies*. Amsterdam: Elsevier Science.

Fonseca, Lois and Jeff Tayman. 1989. "Postcensal Estimates of Household Income Distributions." *Demography* 26:149-159.

Gould, Jeffrey B., Beate Herrchen, Tanya Pham, Stephan Bera, and Claire Brindis. 1998. "Small-Area Analysis: Targeting High-Risk Areas for Adolescent Pregnancy Prevention Programs." *Family Planning Perspectives*. 30: 173-176.

Hamilton, C. Horace and Josef Perry. 1962. "A Short Method for Projecting Population by Age from One Decennial Census to Another." *Social Forces* 41:163-170.

Hashimoto, Shuji, Yoshataka Murakami, Kiyosu Taniguchi, and Masaki Nagai. 2000. "Detection of Epidemics in their Early Stage through Infectious Disease Surveillance." *International Journal of Epidemiology* 29: 905-910.

Longva, Svein, Ib Thomsen, and Paul Inge Severeide. 1998. "Reducing Costs of Censuses in Norway Through Use of Administrative Registers." *International Statistical Review* 66: 223-234.

Morrison, Peter A. and Allan F. Abrahamse. 1996. "Applying Demographic Analysis to Store Site Selection." *Population Research and Policy Review* 15: 479-489.

Murdock, Steve H. and David R. Ellis. 1991. *Applied Demography*. Boulder: Westview Press.

Murdock, Steven H., and Rita R. Hamm. 1997. "A Demographic Analysis of the Market for a Long-Term Care Facility: A Case Study in Applied Demography." In *Demographics: A Casebook for Business and Government*, ed. Hallie J. Kintner, Thomas W. Merrick, Peter A. Morrison, and Paul R. Voss. Santa Monica, CA: RAND.

Rao, J.N.K. 1999. "Some Recent Advances in Model-Based Small Area Estimation." *Survey Methodology* 25: 175-186.

Redfern, Philip. 1989. "Population Registers: Some Administrative and Statistical Pros and Cons." *Journal of the Royal Statistical Society, Series A* 152: 1-41.

Siegel, Jacob S. 2001. *Applied Demography: Applications to Business, Government, Law and Public Policy*. San Diego: Academic Press.

Smith, Stanley K., Jeff Tayman, and David A Swanson. 2001. *State and Local Population Projections: Methodology and Analysis*. New York: Kluwer Academic/Plenum Publishers.

Tayman, Jeff, Bob Parrott, and Sue Carnevale. 1997. "Locating Fire Station Sites: The Response Time Component." In *Demographics: A Casebook for Business and Government*, ed. Hallie J. Kintner, Thomas W. Merrick, Peter A. Morrison, and Paul R. Voss. Santa Monica, CA: RAND.

Thomas, Richard K. 1997. "Using Demographic Analysis in Health Services Planning: A Case Study in Obstetrical Services." In *Demographics: A Casebook for Business and Government*, ed.

Hallie J. Kintner, Thomas W. Merrick, Peter A. Morrison, and Paul R. Voss. Santa Monica, CA: RAND.