5/25/00 admnacs.doc

MEETING STATE AND COMMUNITY NEEDS FOR SOCIAL, ECONOMIC, AND HOUSING INFORMATION:

The Why, What, and How of Converting Program Records and Summarized Survey Data to State and Community Information Systems

Cynthia M. Taeuber
Jacob France Center, University of Baltimore
Bureau of the Census

Julia Lane
Urban Institute

David Stevens
Jacob France Center, University of Baltimore

Background prepared paper for:

Conference on "Developing Public Policy Applications with Summarized Survey Data and Community Administrative Records"
June 6-7, 2000

MEETING STATE AND COMMUNITY NEEDS FOR SOCIAL, ECONOMIC, AND HOUSING INFORMATION:

The Why, What, and How of Converting Program Records and Summarized Survey Data to State and Community Information Systems

Cynthia M. Taeuber
University of Baltimore
Census Bureau

Julia Lane
Urban Institute

David Stevens
University of Baltimore

Abstract

This paper addresses the why, what, and how of creating modern state and community information systems, that is, multiple data sets that are geographically based, to better inform those who make decisions about program effectiveness and direction. Some states and communities have developed their Management Information Systems (MIS) for secondary use as a longitudinal data base for statistical analyses of the overall characteristics of groups of clients, the effects of client groups on other public services, and to evaluate the effectiveness (performance) of program policies. A new information resource profiles the population and housing of communities. That is the American Community Survey. It is a new survey the Census Bureau is developing to provide demographic, social, economic, and housing profiles of areas every year, eventually down to the neighborhood level.

The dynamic picture multiple information sources provide can help communities better understand the interactions between changes in a community's population and other factors such as its industrial mix, economic development potential, health profile, and physical environment. With multiple sources of information, communities can move beyond program administration and performance measures to strategic decisionmaking. Communication of research results to those who make decisions about community policies can improve through new technology such as the layered mapping of Geographic Information Systems (GIS).

This paper describes administrative records as a source for statistical analyses, the new American Community Survey, and other information sources that describe a community's situation compared with other areas. Using current information about population and housing trends from the American Community Survey, we illustrate how administrative records could be merged in econometric models with these summarized survey data to provide improved estimates and probability statements of events. This system protects individual privacy by using data sets matched to small geographic levels rather than individual people. There are other useful additions to econometric models such as the Economic Censuses to describe the economic conditions of metropolitan areas and regions. We address data quality issues in the various data sets and what needs to be done to convert program records to files that are useful for statistical analyses of communities.

Introduction

Need for current information at the community level to track change

Primary responsibility for government program strategies and results has shifted from the federal level to state, county, and community governments in recent years. At a minimum, state and local governments are responsible for reporting measures of

performance to indicate the results of programs paid for with federal tax dollars (Government Performance Requirements Act of 1996). A report from the Committee on National Statistics of the National Academy of Sciences stated: "The devolution of program responsibility …has led to wide variation in programs across states and within states….[this] imposes the need for a significant data infrastructure … [that would] capture state variations in policies and outcomes…." [Moffitt and Ver Ploeg, p. 2]. Further, local governments say that federal reporting requirements are the beginning, not the end – they want solid information they can use for strategic decisionmaking.

We have a system of statistics for the nation as a whole

The United States has a statistical system that provides current information about demographic, social, economic, health, and housing trends for the nation as a whole. This is a system of information based on aggregated data from the decennial census and surveys with a nationally representative design. Census and survey data are collected through a decentralized statistical system of federal agencies coordinated by the Statistical Policy Office of the Office of Management and Budget (OMB). There is a concerted effort by OMB to consult the Congress and other federal policy makers about their information needs and to identify the parts of the data collection system that can most efficiently address national policy questions. Additionally, for statistical purposes only, a very limited group of researchers, who are subject to severe legal restrictions that protect confidentiality, have access to computerized records of individuals who participate in some federal programs.

The national statistical system is designed to address *federal* concerns and meet the requirements of federal laws and court cases. National surveys follow, at best, only general population and economic trends for states. States and sub-states areas can use publicly available data, but the system is designed first and foremost to meet federal data needs to administer programs and distribute funding.

There is no overall system of statistics for states or communities

For geographic subunits of the nation, there is no coordinated "system" to consider overall information needs at the community level or how to meet them. The decennial census, the intercensal population estimates program, and other data sets (usually at the county level) assembled by various federal agencies are the only means the national statistical system provides for examining population, economic, and housing trends among and within communities.

Community policymakers sometimes compensate by using out-of-date census results and current averages from national surveys as ratios to estimate change in their area. Even when the nation is said to be "doing well" economically, there are parts of the country that are not doing well. The extent and distribution of trouble spots is not easy to determine from the statistical system, however. Some liken the situation to putting one's head in a freezer and feet in a fire and then saying, "on average, I feel just fine." The problem is how to determine where the fire is and where the freezer is – that is, which

areas need what and when. Communities want to be proactive rather than reactive, to prevent situations rather than try to mop up after the fact. Without an information system that is current, of high quality, and comparable across areas, it is difficult to adjust policies in time to meet needs, establish priorities, and evaluate results.

New opportunities – data and technology

There are new opportunities to develop community information systems with multiple data sets that provide current and comparable information for models, as well as new software that better organizes data for analysis. This can provide research that better informs those making federal, state, and community policies.

With enhanced models, new software, and the new data sets that are becoming available, we are beginning to have the components for the type of information system at the community level that a modern nation demands. We have the potential to go far beyond traditional uses of decennial census data, to a system that uses multivariate statistical techniques with multiple sources of information to provide improved current descriptions of geographic units as well as predictive models. Time series analyses can be useful in evaluating the impact of a policy change. Researchers have a powerful tool in Geographic Information Systems (GIS) for communicating the results of their analyses to state and community policymakers through the layered mapping of sets of information.

Basic concept of a state and community information system

The basic concepts and a flowchart showing an enhanced state and community information system are shown in Figure 1. The enhanced system is a suite of tools that supplements community administrative records with community profiles based on summarized survey data in models (such as econometric or needs assessment models) to inform policy questions. Among the tools is software that would help community groups with limited resources. For example, such software might organize data for analysis and include automated statistical techniques. Another tool would be improved, inexpensive media for data dissemination (the internet, CD-ROMs, DVDs) and mapping the data.

A system of community information would track the direction of population and housing and compare situations among areas across the nation. It would be able to "generate a profile of short- and long-term outcomes" of programs, produce information about population subgroups at risk of requiring assistance, the duration of episodes, and improve our understanding of how the economic environment affects program success [David, pg. 212; Culhane and Metraux, pp. 345-6].

Such a system would use distributions of the data that are geographically based and aggregated. Rather than relying solely on matching the records of individual people across multiple data files, models would also use aggregated population distributions from different data sets that are matched for small geographic areas. Matching aggregated data at the small geographic level resolves the dilemma of threats to individual confidentiality, as records are anonymous and used at the group level rather

than the individual level. Because access to the confidential records of individuals would no longer be the only means of sophisticated data analyses, more researchers could contribute to the needs of communities for information.

To take advantage of the new data sets and technology, we need to adjust our paradigms for the collection, organization, dissemination, and analysis of the data. And, we need strong laws and protocols that include coordinated statistical policies and using an array of statistical techniques for protecting the confidentiality of information about individual people.

Figure 1. Basic Concepts

WHAT ARE THE POLICY QUESTIONS?

- What is the pattern of relationships among key variables?
- What is the potential future pattern of relationships among key variables?
- What are the results of policy decisions? Can we identify unexpected consequences?

MERGED INFORMATION

- Match to small geography, not individual people, to maintain confidentiality.
- Use summarized profiles from multiple data sets, as appropriate, such as from:

  - ✓ The American Community Survey of the demographic, social, economic, and housing characteristics of areas and changes in areas over time;
  - ✓ Census 2000 and operational tests;
  - ✓ Business and economic profiles from economic censuses and surveys;
  - ✓ Other nationally available, comparable data sets on specific topics;
  - ✓ Multiple sources of administrative records on population groups receiving benefits and services (e.g., food stamps; welfare; subsidies such as transportation, housing, and childcare; unemployment); and
  - ✓ Site locations and information about an area's physical environment (e.g., schools, childcare facilities, public transportation routes, toxic waste dumps, abandoned housing, high crime spots).

MODELS

Develop econometric, needs assessment, and other types of models designed to inform specific policy questions. Establish procedures to ensure that models are evaluated and updated as appropriate. Establish opportunities for coordination and sharing of methods across fields of interest.

ANALYZE IMPLICATIONS OF MODEL RESULTS FOR POLICY

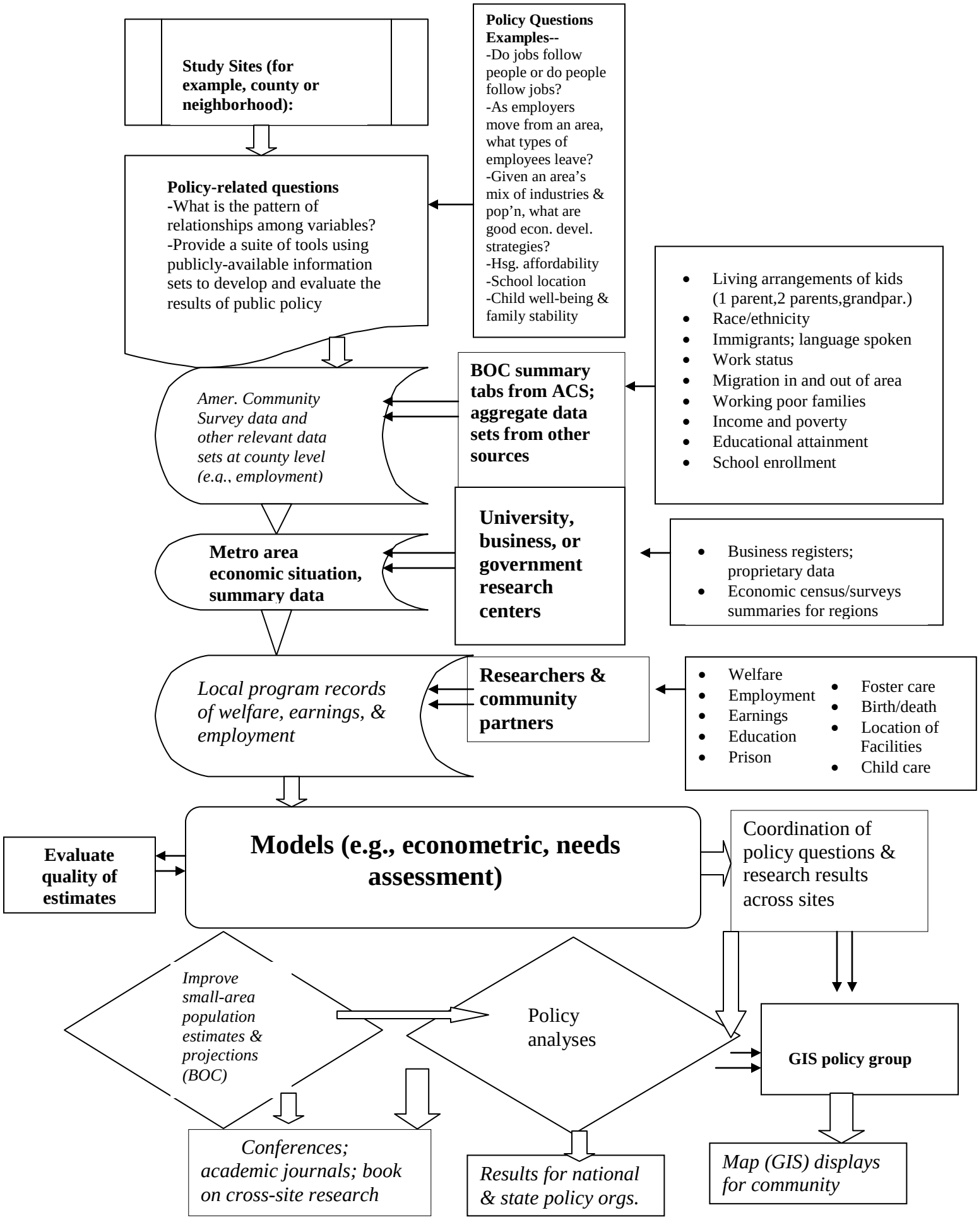
Establish research for specific areas and topics as well as research that reviews and summarizes across individual studies (that is, micro- and macro-level research).

COMMUNICATE RESULTS TO POLICY MAKERS

Establish clear, simple communication of results so they are useful to and used by community groups as well as policy makers at all levels of government, business, and nonprofit organizations.

- Provide a picture of the results through computerized, layered mapping (GIS).
- Automated text and charts in plain language for quick, regular distribution of results.

Figure 1 (continued): System Flowchart



An Enhanced State and Community Information System

There are various data sets that can be used as part of an enhanced state and community information system. These include:

- National surveys that produce limited state and sub-state data;
- decennial censuses;
- the American Community Survey the Census Bureau is developing to replace the decennial census long form;
- surveys state or local governments conduct themselves (produces data that cannot be compared with other areas);
- data sets collected by federal agencies on particular topics such as health or crime;
- estimate programs such as for population counts, poverty, and unemployment;
- lists of physical attributes and the location of facilities; and
- administrative records collected for the management of programs to provide services.

We describe aspects of these data products below. Characteristics of illustrative data sets are shown in Figure 2. Figure 3 compares surveys and administrative records.

Surveys: The American Community Survey and Census 2000

The American Community Survey is being developed by the Census Bureau to replace, the decennial census "long form," that is, the demographic, social, economic, and housing information required by hundreds of federal laws and court cases. The data collection by the American Community Survey will occur throughout the decade rather than just once in ten years. Thus, eventually, the American Community Survey will provide estimates, updated *every year*, of the distribution of *characteristics* of the population and housing in small areas (such as census tracts).

Between censuses, the American Community Survey will improve current estimates of the *number* of people in small areas (such as census tracts and school districts) by furnishing current demographic distributions for use in the population estimate models. What communities have not had previously to inform policy issues and evaluate results, but what the American Community Survey will provide, are estimates of population and housing *characteristics,* and measurements of the level and direction of *change* among areas.

In addition to updated estimates for small geographic areas, the survey will provide estimates for *small population groups* in states and regions. Profiles for small population groups, such as specific Asian and Hispanic nationality groups, children under age 5, people 85 and older, and teenage mothers and whether they are in school or working or unemployed, has previously been available only from the decennial census. With the American Community Survey, we will be able to track trends and the direction of change for population groups. Development of the American Community Survey was started in 1996, and since 1999, is being conducted in 31 diverse sites across the country. The Census Bureau plans to fully implement the survey in every county in the nation

beginning in 2003. Beginning in 2004, and every year thereafter, the American Community Survey plans to have population and housing profiles for areas and population groups of 65,000 or more. For smaller areas and population groups, it will take 3 to 5 years to accumulate information to provide accurate estimates. Information for areas of 20,000 or more people will be available (the 3-year average for the period 2003-05) starting in 2006 and will be updated every year thereafter. Information for areas below 20,000 in population first becomes available in 2008 (the 5-year average for years 2003-2007). After that, the United States will finally have a community information resource that shows change over time, even for neighborhoods and America's small rural areas. Detailed information about the American Community Survey is on the Census Bureau's web site at www.census.gov . American Community Survey data releases to date are available through the American FactFinder at the Census Bureau's web site. Select "Subjects A-Z," then "A," and choose "American FactFinder." A free CD-ROM is available with multiple years of data for sites where data has been collected since 1996 (request by calling 1-888-456-7215 or by e-mail at www.census.gov/acs/www/) .

The data products from Census 2000 (see Attachment B) will be essential resources for state and community researchers while they are waiting for full implementation of the American Community Survey. The first data the census releases tell us about the number of people in geographic areas. By December 31, 2000, the Census Bureau will release the count of the population of states. By April 1, 2001, data users will have counts of the total population and the population 18 years and older tabulated by 63 categories of race – six single race groups and 57 combinations of race.

Throughout 2001 and 2002, various Census 2000 products will be released that provide us with progressively more detailed information about the population and housing characteristics of areas. As part of the decennial census operations, the Census Bureau conducts tests and evaluations and eventually releases information from those studies. One test of interest involves a small national sample that collects information that is essentially the same as the demographic, social, economic, and housing questions on the census long form. This sample provides the Census Bureau an early look at long-form results and essential information about its operational feasibility as they prepare to convert from the long form to the American Community Survey. The data, which will start to become available in mid-2001, is for states, large counties, and the remaining area within each state. Researchers will be able to use this Census 2000 test data as they develop the attributes of the policy-specific models they will use once the American Community Survey is fully operational later in the decade.

Figure 2. Characteristics of Illustrative Data Sets

Data set	Surveys			Model-based estimates	Admin. Records		Physical address of locations; attributes
	Nationally comparable results		Results unique to area		Nationally comparable items	Items unique to area	
	State	Sub-State					
Current Pop. Survey	3-year avg for most	*NYC, LA-Long Beach metro*					No
Am. Hsg. Survey	No	Selected metro areas					No
Decennial census	Yes, every 10 years	Yes, most demogr to block; socio-econ to BG					No
Amer. Community Survey	Yes, 2004+	Yes[1]					No
State-sponsored survey			Current; small sample; high nonresp.				
Population estimates				Including counties, MCDs, incorp. places			No
Local Area Unemploy-Ment Statistics				Incl. metro, nonmetro labor markets, counties, med.cities			No
Poverty Estimates				Counties, school districts			No
School bldg.					No	Yes	Yes
AFDC/ TANF					Yes	Yes	Yes[2]
ES-202					Yes	Yes	Yes[2]
New hires					Yes	Yes	Yes[2]
Crime, type					Yes	Yes	Yes[2]
Health conditions					Yes	Yes	No
Abandoned housing					No	Yes	Yes

[1] Beginning in 2004, and every year thereafter, areas and population groups of 65,000+ will have annual averages. Beginning in 2006, places of 20,000 – 65,000 will have 3-year average estimates updated every year; for areas of 20,000 or less, updated 5-year averages will start to be available in 2008.

[2] Available to researchers only under the rules of confidentiality and laws or agency practices.

Other surveys

As indicated in Figure 2, some national surveys provide sub-national data. The American Housing Survey, for example, includes selected metropolitan areas. The sample size of the Current Population Survey (CPS) is sufficient for annual estimates of the characteristics of the population at the national level and for the largest states. For most states, however, the CPS sample is not large enough to provide annual estimates for their state. Rather, they use 3-year averages from the CPS that are updated each year for characteristics such as income and poverty, educational attainment, patterns of migration, and age distributions. The margin of error is often unacceptably large for measuring year-to-year change such as poverty rates of children in the smallest states.

Some states and communities do their own surveys [Moffitt and Ver Ploeg, p.2]. Advantages include obtaining information that is current and for the total population of the community defined by the survey design. Disadvantages are significant. Surveys are expensive and the results cannot be compared with other areas. Most community surveys have a relatively small sample. Thus, such surveys can provide information only for the total population and broad demographic groups (e.g., "65 and older" rather than information about those with different needs and resources such as the age groups 65-74, 75-84, and 85+). Because classifications are broad and the sample is generally not large enough to include explanatory characteristics (such as teenage mothers and whether they are in school or working), it can be difficult to identify trends that would direct policy choices.

One common way to reduce the cost of state and community surveys is to use untested questionnaires and employ a random digit dialing (RDD) design. RDD surveys are a less expensive way to conduct surveys by telephone than designs that require a field staff and development of a complete list of current households. People without telephones are not included in RDD surveys even though they are often poor and the ones many public policy programs are often meant to serve. Some conduct such surveys in English only, another potential bias depending on the area and the programs for which the information is to be used. Such surveys tend to have high nonresponse rates, a potentially serious data quality issue.

Other nationally comparable data sets

Other nationally comparable data sets are produced by federal agencies for geographic areas (states, counties, and sometimes communities) that researchers can use. For example, at the county level, researchers can use summarized information to describe patterns of school enrollment, income and poverty, employment and unemployment, per capita personal income, sources of earnings, Social Security recipiency, births and deaths, the environment, the number and types of crimes reported, transportation projects, and the incidence of health conditions (for example, low birth weight, substance abuse, and cancer mortality). A useful web site for finding many nationally-available information sets is www.fedstats.gov

Administrative records

Administrative records that are geographically based and result from the administration of federal, state, and local programs have enormous promise as part of an enhanced state and community information system [Stevens, 2000]. Figure 3 compares surveys and administrative records.

Administrative records are created for the management of a program, not for statistical analyses. Often, variables are not comparable among areas, especially across state lines. Nevertheless, states and communities have begun to realize and act on their potential for current, albeit incomplete information of varying quality, about segments of the population [for example, see Culhane and Metraux 1997; Mueser, et. al., 1999]. For example, there are records about an area's housing, such as age of the house and property taxes. Such tracking systems provide longitudinal information that researchers can use to calculate prevalence rates, determine the characteristics of people who use services over time, and measure the effectiveness of various programs and how changes in one program affect the use of other programs [Culhane and Metraux, 1997, p. 343]. As Martin David has pointed out, the use of multiple sets of administrative records would better inform us about "how government activities in several agencies relate to each other, what government services cost, what groups are covered by benefits, and what groups are excluded from benefits" [pg. 214].

Gaining access to publicly-available survey results is becoming easier every year. As Culhane and Metraux say in a mild understatement, "Gaining access to data from other agencies is often fraught with difficulties…and in many cases will be impossible." [p. 351]. The legal restrictions and compatibility issues are obvious barriers. State agency heads determine whether the administrative records of their agency can be used. They sometimes see research as a "report card" that may cause them more problems than benefits.

As we discuss in the sections that follow, by increasing the ability to use multiple sources of information with confidentiality protocols, the nation could realize both a community information system and research that is useful for planning and evaluation of programs. The significant benefit of expanded research opportunities to answer policy questions is counterbalanced by concerns for and the need to develop strict protocols, both legal and technical, to protect the confidentiality and safety of the individuals within the data systems. Along with addressing confidentiality concerns, there are technical steps that should be taken to make program files useful and of acceptable quality for statistical analyses. This ideally includes, for example, standardization of definitions across files and protocols for handling missing and conflicting information.

Figure 3. Comparison of Surveys and Administrative Records

Characteristics	Surveys w/substate data	Administrative records (AR)
Currency of data	Release varies (e.g., 6 months after data collection year for the American Community Survey but 2 years for decennial census long form)	Release varies but generally within a year
Time frame of survey design	Usually cross-sectional	Usually longitudinal
Universe	Total population for federally sponsored surveys. State/local sponsored often telephone surveys to reduce cost (bias of not including people without phones; English language only; nonresponse often high).	Program participants; no information about remainder of population of area. As compared with surveys, greater risk of double counting. Does not include those eligible but not receiving services.
Confidentiality protocols	Public-use summary and sample of micro-data files regularly prepared. Access to full micro data files strictly limited. Social Security Number (SSN) generally not collected.	Varies according to state laws and practices. Profiles for small areas not generally available or comparable across areas. SSN usually available as identifier across files.
Margin of error from sampling	Increases the smaller the geographic area; often must use multi-year averages to obtain reliable data. May be unacceptably large for measuring year-to-year change.	Not applicable as all members of subgroup receiving services are represented.
Comparability across areas	Federal surveys comparable across areas. State surveys unique to area.	Only a few nationally comparable AR files (e.g., food stamps; AFDC/TANF)
Geo-coding of addresses	Census Bureau provides consistent sampling frame and geocoding for most federal surveys	Conventions and quality vary although this is key to correct assignment of characteristics to specific areas.

Data collection and processing procedures	• Interviewers formally trained on questionnaire • Information usually from self report of respondent • Statistical techniques used for item nonresponse and inconsistent information • Standards for data entry errors	• Interviewers trained to collect information for program management, not statistics • Information usually observation of interviewer mixed with self reports • Nonresponse and inconsistencies in data set • Definitional differences across data sets significant
Detail of information	Usually broad profiles	Rich detail on topic of data set, Usually includes dates, duration of service use, some personal characteristics of clients.
Cost	Relatively expensive as compared with administrative records	Data required for program management; cost in converting to statistical files (e.g., geocoding addresses, edits, software)

Filling data gaps with both surveys and administrative records

The great promise for developing a statistical system for states and communities lies in taking advantage of the characteristics of administrative records and surveys. Together, surveys and administrative records fill gaps that cannot be met when researchers can use only one or the other. For example, school enrollment records provide information about students aged 5 to 18 years. School officials have a difficult time projecting where to locate future schools because they have no current information about the numbers or characteristics of the population under age 5 or the age ranges or the demographic and economic characteristics of children moving in and out of the school district. That information will be provided by the American Community Survey, as described above.

Using administrative records with survey information that can be compared among areas can be useful for both state and federal policy decisions about funding allocations and the direction of policy. For example, the Jacob France Center at the University of Baltimore is working with a consortium of university research centers in six urban areas to analyze welfare-related transitions. Each center uses confidential administrative records (such as welfare, employment, and earnings records) that have been converted to use as statistical files for use in econometric models to inform policy. Once they can add current population trends from the American Community Survey, the researchers can incorporate the current demographic distributions into their econometric models and hereby broaden the types of questions they can address. For example, as welfare caseloads drop nationally, they can better consider the heterogeneity among areas to judge what types of areas follow national trends and which do not. They can ask questions such as, "Are the working poor and welfare recipients more likely to combine households to cope with their economic situations and is that correlated with changes in the welfare system?" or "Do housing and transportation subsidies help the working poor remain in the workforce?"

Another use of multiple data sets is to obtain information at the community level that would be unreasonably expensive if collected by a survey alone. For example, HUD has initiatives to encourage inner city economic development. Likewise, the Department of Agriculture has programs to encourage economic development in rural areas. The decisionmakers they work with to meet the objectives of the programs are frustrated by the lack of information about consumer expenditure patterns to describe community buying power. Standard sales forecasting models do not work for inner city markets, for example, and need to be adapted to show whether a neighborhood can support retail stores such as grocery stores and pharmacies. The information that would help is similar to that found in a national-level Consumer Expenditure Survey. Replicating it at the community level would be a prohibitive cost and burden on the public. It is conceivable, however, that indicators from the American Community Survey (e.g., age distributions, educational attainment, home ownership, and employment) in combination with data businesses collect (such as sales in a similar area), could provide a sufficiently reliable

model of buying power to inform the strategy of those seeking to develop the economy of a particular inner city.

Traditionally, the input for models is a record of information for individual people collected from one or more sets of administrative records. While it is cost-efficient for researchers to have access to individual records to evaluate interaction effects among programs, American society places high value on protection of individual confidentiality and our legal system reflects that. For example, sworn Census employees (including with Special Sworn Status) are subject to fines up to $5,000 and five years in prison if they reveal any individual information from the American Community Survey, the decennial census, the Economic Censuses, and other surveys protected by Title 13 of the U.S. Code. The confidentiality of individual information in administrative records is also protected under various federal and state laws. This severely limits the number of researchers who can use these sources for complex questions.

There is a substantial new tool for providing an enhanced information system for states and communities. It is supplementing the information from administrative records (whether at the unit level or aggregated) with aggregated summaries derived from publicly available survey data for small geographic areas as input for models. Under this method, it is not necessary to match individuals across files. Therefore, many researchers, rather than only a few, can use aggregated summaries to work on questions important to communities. This approach does not answer all questions, but it will afford researchers new opportunities to understand the interactive effects of events and for policymakers to consider preventive actions. A second expectation is that such research goes beyond the traditional methods of historical descriptions to significantly improve our ability to predict the course of population and economic events through probabilistic modeling.

A system that uses multiple data sets furnishes insight into the interactions between changes in an area's population and industrial mix, and other factors such as health conditions [Craney, Padgett, and Lorimer], service systems [Culhane and Metraux, p. 342], and the environment. Examples from federal programs include the Census Bureau's Small Area Income and Poverty Estimates Program (SAIPE) and program to make population estimates between censuses, and the Local Area Unemployment Statistics (LAUS) program of the Bureau of Labor Statistics (see Attachment A for more information on both programs). Basically, the concept is to merge aggregated data systems on population and housing characteristics, metropolitan-area economic situations, and locally available computerized records on clients who use various community programs into models to answer policy questions (see Figure 1). This methodology allows predictive modeling and provides estimates of population characteristics that are an improvement over those available from any one data system alone. The results can help meet community needs for social, economic, and housing information for making policy decisions and evaluating program progress and direction. The improved estimates and methods can help communities make projections of their outlook that could help them as they look towards investments in their future.

The models do need to be benchmarked and updated. In particular, it is important to compare the summary information derived from the merged aggregate data with that which can be derived from merged micro data. How might this be done? Such micro data exist at the Census Bureau, but the sensitive nature of the data dictates that access is extremely restricted – the data are within a firewall behind a firewall behind a firewall. However, validation of models is certainly feasible for approved Title 13 projects and researchers with Special Sworn Status at Census Bureau headquarters in Suitland, MD.

Along with modified models, new software that better organizes data for analysis is rapidly being developed. There is need for easier access to data and improved displays of research results. For example, a difficult issue is how to help decisionmakers understand the limits of the data they are using by including information about the size of sample errors from surveys in reports, charts, and communication media such as GIS. The American Community Survey reports the 90-percent confidence interval in its data products to alert data users to the reliability provided by the sample size for the subject matter detail and geographic area of interest. Leitner and Buttenfield [2000] suggest graphical guidelines that should be incorporated as defaults in GIS software.

Another need is for a "more accurate nationwide shared geographic database that can alternately use both address and geographic coordinate as entry points or match keys …[because of] the increasing need to integrate environmental, health, and population databases at local and regional and state levels…" [Sperling and Sharp, p. 38]

One powerful new tool is the ability to overlay maps of data on different topics with the software, Geographic Information Systems (GIS). Nancy LaVigne (National Institute of Justice) has pointed out that GIS can serve as a data integrator, an analysis tool, and a presentation vehicle. She says, "As a data integrator, GIS enables researchers to merge community-level data from disparate sources based on common geographic references. The analysis capabilities of GIS offer a powerful means of examining spatial clustering, correlation, and adjacency effects. In addition to these analysis functions, the power of the visual display of data in map form makes GIS an ideal tool for communicating research findings to policymakers and practitioners" (LaVigne, June 7, 2000]. Culhane, Lee, and Wachter, for example, used multivariate analyses to examine housing and neighborhood conditions, such as housing code violations and crime, in relation to the last home addresses of people who had subsequently become homeless. Dynamic models of neighborhood change (population characteristics, housing stock, physical condition) could inform those designing, targeting, and siting prevention programs or suggest whether other strategies, such as targeting high crime areas to reduce residential instability, would be effective [Culhane, Lee, and Song].

Examples of the types of policy questions communities could answer with multiple data sets merged in econometric models:

Generally, a community information system could provide:

- Time series data for analyzing community trends.
- Estimates of the size and composition of population groups.
- Evaluations of the effectiveness of programs and policies.
- Improved understanding of the effect of population group dynamics on service systems (public and private) and related service systems (for example, the effect of welfare leavers on service systems for the homeless population, and on the health care and justice systems)
- Predictors of events, population changes, prevalence, and relative risks of events.
- Neighborhood and other small-area indicators.

Some specific examples of the types of policy questions communities could answer with multiple data sets merged in econometric models are shown below:

- Economic development:

  - ✓ Which population groups benefit the most from the types of jobs being created (or lost) in an area? Who are the winners and losers?

  - ✓ To what extent do jobs follow people and people follow jobs? Causality?

  - ✓ As employers move from an area, what types of employees leave? Stay? Likewise, when employers move into an area or expand, what population groups are drawn into the area?

  - ✓ Given an area's mix of industries and population, what are the options for economic development strategies? What types of industries are hiring high school dropouts, how well do the jobs pay and what is the duration of such employment? What happens as unemployment levels go up (down)?

  - ✓ Population effects are interrelated with the business conditions in the area, but what are the interrelationships? Because of confidentiality and competitive business interests, economic data about businesses are presented only at high geographic levels or as basic information for smaller geographic areas such as that from County Business Patterns. Additionally, population and business censuses have been conducted on different schedules. The coming availability of yearly population summaries from the American Community Survey will allow us to look at population and business effects in time periods that coincide. Econometric models will be able to provide predictions of population-business interactions for smaller areas based on the relative population and business conditions within a region.

- Transitions in welfare to work:

  - ✓ As caseloads drop nationally, what is happening at the local level? Demonstrate the heterogeneity among sites in changes in local economic conditions and caseloads. What types of areas follow national trends and which do not?

  - ✓ What has happened to those who have left the welfare system but who are not working in the area? Where are they (e.g., in school, in prison, homeless, moved from area and working elsewhere)? What are the patterns among areas?

  - ✓ Is there resource sharing? Are there changing household demographics among the poor with children correlated with changes in the welfare system? For example, are more households being combined among welfare recipients than among the working poor?

  - ✓ What population groups are eligible for program services but not enrolled as expected? What is the relationship between changes in poverty rates and caseloads? How does this differ among types of areas?

  - ✓ What supports are needed to keep low-wage workers on the job? Are transportation and housing subsidies needed to encourage entry and remaining in the workforce at relatively low wages? What level of cost can low-wage workers afford themselves? What are the tradeoffs between subsidies for housing versus transportation to reduce the effect of the mismatch between residence and job location of low-wage workers?

  - ✓ Educational outcomes -- what is the earnings stream of those who have vocational training or community college? Does this differ among types of areas?

  - ✓ Child well-being – What is their stability of residence and household composition? Childcare? Foster care?

  - ✓ Youth – What are the school-work-welfare transitions among adolescents and how do they differ among different types of areas?

  - ✓ What are the differential effects of barriers such as prison records, lack of education, and use of a language other than English to employment and income among different demographic groups (such as age, marital status, race/ethnicity)?

- Health needs assessments:

  - ✓ How do changes in population and industry mix affect health needs and ability to pay?

  - ✓ What can you predict about future needs given an area's migration patterns and changes in demographic composition?

- ✓ What is the impact of different environmental conditions on health needs among different demographic groups (e.g., race/ethnic groups, age, economic status)?

- Projecting future needs for community facilities such as schools and hospitals:

  - ✓ Use population surveys to determine migration patterns, demographic group (e.g., age, gender, economic status, educational attainment) with school enrollment records (or hospital use records) to project future needs for location of facilities and types of facilities (such as temporary trailers or long-term school buildings; geriatric care or prenatal care facilities).

- Neighborhood development and housing policy:

  - ✓ What is the likely effect on the affordability of housing and the composition of the population?

  - ✓ What are the socio-economic effects of living in particular types of neighborhoods relative to family types? What is the duration of residence in neighborhoods with concentrated poverty? How does residence in such areas affect employment opportunities?

- Safety:

  - ✓ How does the incidence of crimes (by type of crime) change in neighborhoods with changes in migration patterns into and out of those neighborhoods and changes in the local economy?

- Energy and the environment:

  - ✓ What are the predicted changes in energy demand or environmental quality as the mix of population and industry changes?

Example of How Multiple Information Sources Might Be Used

How could a county economic development council, in the year 2000, approach policy questions using existing multiple data sources in an econometric model? The Census Bureau, in preparation for the full implementation in 2003, is asking for feedback on which tabulations best respond to public policy questions. This provides a window of opportunity for the research community to design inputs into the decisionmaking process. In order to illustrate an example of how this might be done, we pose the following illustrative policy question:

Given a county's mix of industries and population, what are some options for economic development strategies that focus on improving the situation for the county's low skill

workers? What types of industries are hiring low skill workers, and how well do the jobs pay? What happens as unemployment levels go up (or down)? What if there is a downturn in the economy?

Although these are core economic development issues, the data available for describing these outcomes econometrically are scarce and not available in any one data set. Ideally, one might want to look at a model that either structurally describes the demand and supply for low skilled workers, or sets up a reduced form model of the type:

$$\text{Employment}_{\text{low skill, industry, time}} = X\beta + Z\delta + CC\,\lambda + \varepsilon$$
$$\text{Earnings}_{\text{low skill, industry, time}} = X\,\alpha + Z\phi + CC\,\varphi + \mu$$

where:

X reflects the characteristics of the available labor pool of workers (such as age, race, sex, education, prior work history, welfare recipiency),

Z reflects the characteristics of the firms in each industry (such as the number and average size of firms, total employment in the industry, average payroll, turnover rates, and growth rates), and

CC reflects county conditions, such as the local unemployment rate, or statewide economic factors.

Once this model is estimated, the effect of structural demand side changes in the economy can be modeled by simulating different changes in Z. Changing CC can simulate the effect of changing economic conditions.

The first step for every researcher is to find the data sets that are adequate to address the policy issues. A stocktaking of available sources would reveal that there might be several relevant data sets available: the 1997, 1998, and 1999 American Community Survey; the 1990 census; school enrollment records (every year through 1999); AFDC/TANF records; and Unemployment Insurance (UI) wage records.

How might these be used in the simple model described above? The dependent variables - the employment and earnings of low skill workers - could be derived by year and industry from both the 1990 census and each year of the American Community Survey. The workforce characteristics - educational attainment of the population, as well as the prior work history, age, race and sex distributions - could be derived from both the 1990 census as well as each year of the American Community Survey. Administrative records on TANF and AFDC recipiency could be aggregated to derive the proportion of welfare recipients in the population. In addition, if the recipiency were matched to UI data, it could even be used to reveal which types of industries hired welfare recipients, and this information could be included in the vector of Z variables.

Finally, the variables on the firm side, such as the average size of firm, payroll, turnover rates, and growth rates of firms in each industry could be derived from UI wage records and ES-202 files for each quarter of 1990-99.

These can then be used to simulate changes in the demand for low skilled workers as a given industry grows or contracts. If, for example, firms in health services were projected to become an important segment of the economy, the effect of this increased growth on the employment and earnings of unskilled workers could be simulated, assuming, of course, that other things were held constant!

The point of this exercise was to determine what the key tabulations are likely to be. In this case, the key tabulations from the American Community Survey are employment and earnings by type of worker and by industry; age, race, gender, and educational experience, by county. These can be directly inserted into a regression model, together with similar tabulations from administrative data. Although different applications might well be envisaged, these kinds of tabulations would certainly enhance the ability of researchers to model important, policy relevant, questions.

Preparing For A Multiple Choice Future

We have discussed a vision of developing a coordinated state and community information system for strategic decisionmaking that makes efficient use of technology yet honors and protects confidentiality. The potential for such information systems is huge. The ability to do such research could affect, for example, how public programs are "funded, targeted, and evaluated," and would be useful in designing programs for high risk groups and assessing needs and resources [Culhane and Metraux, p. 347].

There are significant barriers, however, that require sustained and coordinated effort to overcome. For example:

- The American Community Survey must be fully implemented nationwide with a sample that is large enough to provide updated estimates of demographic and housing characteristics for census tracts.

- Statistical policy should be coordinated among the multiple data sets of administrative records to standardize, to the extent possible, definitions, ways to ask demographic questions (for example, age and race), processing and editing rules such as for missing or inconsistent data, and the coding of addresses and subjects such as industry and occupation. Where possible, it would be advantageous for the standardization to be consistent with the conventions used in national surveys such as the American Community Survey.

- Data quality issues in administrative records need to be evaluated and addressed. For example, the people who collect information from program participants think of its use in terms of a management information system, not as statistical files for research. Training such staff about the extended uses of the data helps them understand why it

is important to gather the information according to standard practices and to try to obtain all the information requested. We have an example of the effectiveness of this approach from the National Center for Health Statistics. It trains funeral directors about the uses of the information requested on death certificates and has thereby improved the quality of that data.

- Much has been written about the difficult balance between individual rights to privacy and the confidential use of administrative records to achieve efficient, fair government [Stevens; Warren and Brandeis, p. 193; Hatch, p. 1; National Academy of Sciences; Human Rights Commission]. Stevens notes that "privacy" refers to the right of individuals to control information that is about them while "confidentiality" refers to how information is handled and what the statutory responsibilities are [Stevens, p.7]. Protocols are necessary to protect the confidentiality of the physical records as well as avoiding the disclosure of an individual's identity and attributes when information is released in a statistical file.

  - ✓ Confidentiality is protected through laws and regulations, statistical policy, and statistical techniques. It is especially important that laws and regulations distinguish between administrative and statistical uses of information about an individual. Administrative uses explicitly affect an individual, such as providing or denying benefits or identifying illegal activities. Statistical analyses are uninterested in an individual once the person's characteristics are collected and then aggregated with those of others to identify patterns. Current legal language is often open to interpretation as to whether and how third party researchers use administrative records [Stevens, pg. 4].

  - ✓ Statistical policies to protect individual confidentiality may include, for example, control over who has access to individual records, an institutionalized expectation and ethic that confidentiality is everyone's responsibility, a disclosure review board, and the requirement that every staff member sign a statement every six months that (s)he knows and understands the laws and practices to secure information. A significant problem is that, with multiple agencies, executive responsibilities, philosophies, and roles among legal counsel, the lack of an overall policy results in ad hoc and ever changing practices [Stevens, pg. 4]. Uniform data sharing agreements that cover multiple data sets would be useful.

  - ✓ Statistical techniques for disclosure avoidance address issues such as the minimum number of cases required before data are shown for an area, the detail that is provided in cross tabulations in relation to a population size, switching of households among areas so one can never be certain of identity, and techniques that avoid disclosure in data sets from Year 1 to Year 2 when a city annexes a few blocks (that is, information can be gained about the annexed blocks by subtraction). Advanced techniques for disclosure avoidance and disclosure detection are needed for use with Geographic Information Systems (GIS).

- We need to further develop software to make data access easier for staff in community organizations who are not trained demographers or economists. They also need software that better organizes the data for predetermined analyses and statement of results.

- We need to improve the geographic coding of addresses on administrative records and further develop software such as GIS for better communication of research results to community organizations and policymakers.

References

Bollinger, Christopher R. and Martin H. David. 2000. *"I Didn't Tell and I Won't Tell": Dynamic Response Error in the SIPP*. Report for the National Science Foundation.

Craney, James, with Paige Padgett and Ronald Lorimer. 1998. The Level of Community Well-Being and Selected Maternal and Child Health Indicators: The Houston MSA. See website of Healthy Communities Project at the University of Texas-Houston School of Public Health at: <www.sph.uth.tmc.edu/chis/background.htm>

Coulton, Claudia J. 1999. Public Assistance Records: A Source for Neighborhood Indicators. Washington, D.C.: National Neighborhood Indicators Partnership, The Urban Institute.

_____, with Lisa Nelson and Peter Tatian. 1997. Catalog of Administrative Data Sources: For Neighborhood Indicator Systems. Washington, D.C.: National Neighborhood Indicators Partnership, The Urban Institute.

Culhane, Dennis P. and Stephen Metraux. 1997. Where to from Here? A Policy Research Agenda Based on the Analysis of Administrative Data. In *Understanding Homelessness: New Policy and Research Perspectives*, ed. Dennis P. Culhane and Steven P. Hornburg, 341 – 380.

_____, Chang-Moo Lee, and Dayoung Song. 1997. Spatial-Temporal Analysis of Neighborhood Dynamics and Homelessness Incidence: The Effect of Neighborhood Deterioration, Gentrification, and Minority Segregation. Unpublished paper. University of Pennsylvania.

_____, Chang-Moo Lee, and Susan M. Wachter. 1997. Where Homeless Come From: A Study of the Prior Address Distribution of Families Admitted to Public Shelters in New York City and Philadelphia. In *Understanding Homelesseness: New Policy and Research Perspectives*, ed. Dennis Culhane and Steven Hornburg, 225-63. First published as *Homeless Policy Debate*, 7(2):327-65.

David, Martin H. 2000. "Monitoring Income for Social and Economic Development," in Burt S. Barnow, Thomas A. Kaplan, and Robert A. Moffitt (eds.), *Evaluating Comprehensive State Welfare Reforms: The Wisconsin Works Program.* Albany, NY: Rockefeller Institute Press. (Previously published as *Special Report Series 69*, Madison, WI: Institute for Research on Poverty.)

Hatch, Orrin G. 1983. Balancing Freedom of Information With Confidentiality for Law Enforcement, *Journal of Contemporary Law*, V9.

Heeringa, S.G. (1982). "Statistical Models for Small Area Estimation," in *Readings in Population Research Methodology*, 5, 126-132.

Human Rights Commission. 1993. *14th International Data Protection & Privacy Commissioners Conference: Conference Papers.* Sydney, Australia.

Kingsley, G. Thomas (ed.). 1999. *Building and Operating Neighborhood Indicators Systems: A Guidebook.* National Neighborhood Indicators Partership Report. Washington, D.C.: The Urban Institute.

LaVigne, Nancy. 2000. Statement for a roundtable at the June 6-7, 2000 conference, "Developing Public Policy Applications with The American Community Survey and Community Administrative Records," Jacob France Center, University of Baltimore.

Leitner, Michael and Barbara P. Buttenfield. 2000. *Guidelines for the Display of Attribute Certainty.* Cartography and Geographic Information Science, Vol. 27, no. 1. Pp. 3-14.

Moffitt, Robert A. and Michele Ver Ploeg, eds. 1999. *Evaluating Welfare Reform: A Framework and Review of Current Work.* Committee on National Statistics, Commission on Behavioral and Social Sciences and Education, National Research Council. Washington, DC: National Academy Press.

Mueser, P., J. Hotchkiss, C. King, P. Rokicki, and D. Stevens. 2000. The Welfare Caseload, Economic Growth, and Welfare to Work Policies: An Analysis of Five Urban Areas. Report for The Jacob France Center, University of Baltimore.

National Academy of Sciences. 1993. *Private Lives and Public Policies: Confidentiality and Accessibility of Government Statistics.* Washington, D.C.

Olson, J., C. King, J. Hotchkiss, P. Mueser, P. Rokicki, and D. Stevens. 2000. Welfare-to-Work Transitions in Five Urban Areas: Initital Results from the Pooled Multivariate Analysis. Report for The Jacob France Center, University of Baltimore.

Sperling, Jonathan, and Stephen A. Sharp. 1999. *A Prototype Cooperative Effort to Enhance TIGER.* URISA Journal (Summer), pp. 35-42.

Stevens, David W. 2000. Administrative Record Linkage: Promise and Pitfalls. Baltimore, MD: The Jacob France Center. 61 pp.

_____. 1996. Towards an "All Purpose" Confidentiality Agreement: Issues and Proposed Language. Report for the Office of the Assistant Secretary, Maryland Department of Labor, Licensing, and Regulation.

Warren, Samuel D. and Louis D. Brandeis. 1890. The Right to Privacy, *Harvard Law Review*, IV:5 (December).

Attachment A
Examples of Federal Programs for Estimating Distributions from Multiple Data Sets in Models

Small Area Income and Poverty Estimates (SAIPE)

An example of the concept of making estimates for small areas by combining survey data for states with county-level program records is the Census Bureau's program for making estimates of poverty for small areas, the Small Area Income and Poverty Estimates (SAIPE) program. The SAIPE program uses intercensal estimates of the population of states and counties by age and group quarter status. It uses income and poverty estimates from the 1990 census for states and counties and from the Current Population Survey for states (that is, aggregated distributions for each state). These aggregated data are combined with program records available for all counties in the nation in statistical models to produce estimates of poverty for school-age children for counties and school districts. The program records SAIPE uses are the number of food stamp and Social Security recipients, summarized data from tax records, and estimates of personal income. Estimates of poverty and income are made using survey and administrative data that are not available until two years after the year to which they refer. Thus, income year 1997 estimates cannot be produced until the year 2000. Detailed information about the SAIPE program is available on the Census Bureau's web page at:
<www.census.gov/hhes/www/saipe/techdoc>

Population Estimates Between Censuses: Concepts and Geography

The information below is extracted from the Census Bureau's website and it was last revised on March 20, 2000. For more information contact U.S. Census Bureau, Population Division, Population Estimates Branch.

What is a population estimate?

The Census Bureau's Population Estimates Program (PEP) produces July 1 estimates for years after the last published decennial census (1990), as well as for past decades. Existing data series such as births, deaths, federal tax returns, Medicare enrollment, and immigration, are used to update the decennial census base counts. PEP estimates are used in federal funding allocations, in setting the levels of national surveys, and in monitoring recent demographic changes. A methodology reference accompanies most of our population estimates offerings.

Revisions to estimates and geographic detail

With each new issue of July 1 estimates, PEP revises estimates for years back to the last census. Previously released estimates become superseded. Revisions to estimates are usually due to input data updates, changes in methodology, or legal boundary changes.

The frequency of estimates and availability of demographic detail vary by geographic level.

Why does the Census Bureau produce estimates?

The legal requirement for the Census Bureau to produce subnational population estimates is given in Title 13 of the U.S. Code. Title 13 states that: "During the intervals between each census of population required under section 141 of this title, the Secretary, to the extent feasible, shall annually produce and publish for each State, county, and local unit of general purpose government of fifty thousand or more, current data on total population and population characteristics and, to the extent feasible, shall biennially produce and publish for units of general purpose government current data on total population." The reason for producing estimates is given in Section 183 of Title 13: "Except as provided in subsection (b), for the purpose of administering any law of the United States in which population or other population characteristics are used to determine the amount of benefit received by State, county, or local units of general purpose government, the Secretary shall transmit to the President for use by the appropriate departments and agencies of the executive branch the data most recently produced and published under this title."

In other words, the Census Bureau produces subnational estimates for use in the allocation of funds to state, county, and local governments. For this reason, the Census Bureau produces population estimates for general-purpose functioning governments. These governments have elected officials who can provide services and raise revenue. In addition to states and counties, incorporated places and minor civil divisions also serve as general- purpose functioning governmental units.

For what geographic areas does the Census Bureau produce estimates?

In addition to the Nation, the 50 states, and the District of Columbia, PEP produces estimates for the following geographic entities:

Counties (and equivalents)

Counties are the primary legal divisions of most states. Most counties are functioning governmental units, whose powers and functions vary from state to state. In Louisiana, these primary divisions are known as parishes. In Alaska, the county equivalents consist of legally organized boroughs or "census areas" delineated for statistical purposes by the State of Alaska and the Census Bureau (since 1980). In four states (Maryland, Missouri, Nevada, and Virginia), one or more cities are independent of any county organization and thus constitute primary divisions of their states; the Census Bureau refers to these places as "independent cities" and treats them as the equivalents of counties for estimates purposes. The District of Columbia has no primary divisions and the entire area is considered to be the equivalent of a county and in Puerto Rico, municipios are the primary divisions and treated as county equivalents for estimates purposes. Legal changes to county boundaries or names are typically infrequent. Changes that have occurred since the 1990 Census are documented at: http://www.census.gov/population/www/estimates/geonotes/

These notes also include information on changes to the areas described below.

Minor Civil Divisions

Legally defined county subdivisions are referred to as minor civil divisions (MCDs.) MCDs are the primary divisions of a county. They comprise both governmentally functioning entities - - that is, those with elected officials who provide services and raise revenues -- and nonfunctioning entities that exist primarily for administrative purposes, such as election districts. Twenty-eight states and Puerto Rico have MCDs. However, the MCDs function as general purpose governmental units in all or part of only twenty states. Within these twenty states, PEP produces estimates for all governmentally functioning MCDs and for nonfunctioning MCDs in counties that contain at least one functioning MCD.

The legal powers and functions of MCDs vary from state to state. Most of the MCDs in twelve states (Connecticut, Maine, Massachusetts, Michigan, Minnesota, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont, and Wisconsin) serve as general-purpose local governments. In the remaining eight states for which PEP produces MCD level estimates (Illinois, Indiana, Kansas, Missouri, Nebraska, North Dakota, Ohio, and South Dakota) the MCDs, for the most part, perform less of a governmental role and are less well known locally, even though they are active governmental units.

MCDs primarily are known as towns (in New England, New York, and Wisconsin), townships, and districts, but also include a variety of other entities. In Maine and New York, American Indian reservations are not part of any other MCD and therefore, the Census Bureau treats them as MCDs. PEP does not produce separate estimates for American Indian Reservations regardless of their MCD status. In some states, all or some incorporated places are subordinate to the MCDs in which they are located. Therefore, a place may be either independent of or dependent upon MCDs. In one state (Ohio), a multi-county place may be treated differently from county to county. The District of Columbia defined no MCDs for the 1990 census, so the District itself serves as the equivalent of an MCD for data presentation purposes. No functioning MCDs exist in Puerto Rico.

Incorporated Places

The legal designations, powers, and functions of incorporated places vary from state to state. Incorporated places include cities, towns (except in New England, New York, and Wisconsin where the Census Bureau recognizes towns as MCDs for census purposes), boroughs (except in Alaska, where the Census Bureau recognizes boroughs as equivalents of counties, and New York, where the Census Bureau recognizes the five boroughs that constitute New York City as MCDs) and villages. Incorporated places can cross both county and MCD boundaries. When this occurs, the place name is followed by the designation "pt" (which stands for part). The PEP produces estimates of the

unincorporated "balance of county" area for counties that are not entirely composed of incorporated places. Another way to understand this is to think of the "balance of county" as the county population minus the county population resident within incorporated places.

Consolidated Cities

Consolidated cities are a unit of government for which the functions of an incorporated place and its county or MCD have merged. The legal aspects of this action may result in both the primary incorporated place and the county or MCD continuing to exist as legal entities, even though the county or MCD performs few or no governmental functions. Where one or more other incorporated places within the consolidated government continue to function as separate governmental units, the primary incorporated place is referred to as a "consolidated city."

Estimates are not shown for consolidated cities. Rather, estimates are displayed for the consolidated city "remainder," which is the consolidated city minus the semi-independent incorporated places located within the consolidated city. Consolidated cities include: Butte-Silver Bow, MT; Athens-Clark County, GA, Augusta-Richmond County, GA, Columbus, GA; Indianapolis, IN; Jacksonville, FL; Milford, CT; and Nashville-Davidson, TN. Estimates also are produced for the semi-independent places which together with the "remainder record," sums to the entire territory of the consolidated city.

Additional Information on Geographic Entities

A more complete narrative treatment of these areas is found in the Geographic Areas Reference Manual: http://www.census.gov/geo/www/garm.html authored by the Census Bureau's Geography Division. This manual provides a comprehensive description of all the geographic entities recognized and reported in the Census Bureau's various Censuses and Surveys.

Local Area Unemployment Statistics (LAUS) -- Overview

The Bureau of Labor Statistics (BLS) has a Local Area Unemployment Statistics (LAUS) program. Unemployment statistics for local areas are model based and use multiple, nationally available data sets. BLS is conducting research so that once the American Community Survey is fully implemented, they hope to use those current demographic estimates to replace the historical information from the previous decennial census. The information below describes the current LAUS program and is excerpted from the website of the Bureau of Labor Statistics. The address is: www.bls.gov/lauov.htm

The Local Area Unemployment Statistics (LAUS) program is a Federal-State cooperative effort in which monthly estimates of total employment and unemployment are prepared for approximately 6,800 areas:

- Census regions and divisions
- States
- Metropolitan areas (primary metropolitan statistical areas and metropolitan statistical areas)
- Nonmetropolitan labor market areas
- Counties and county equivalents
- Cities of 25,000 population or more
- Cities and towns in New England regardless of population

These estimates are key indicators of local economic conditions. The Bureau of Labor Statistics (BLS) of the U.S. Department of Labor is responsible for the concepts, definitions, technical procedures, validation, and publication of the estimates that State employment security agencies prepare under agreement with BLS.

The concepts and definitions underlying LAUS data come from the Current Population Survey (CPS), the household survey that is the official measure of the labor force for the nation. Annual average data for all States, the District of Columbia, New York City, and the Los Angeles-Long Beach metropolitan area are derived directly from the CPS. Monthly estimates for these areas are produced using estimating equations based on regression techniques. These models combine current and historical data from the CPS, the Current Employment Statistics (CES) program, and State unemployment insurance (UI) systems. Estimates for substate labor market areas (other than the two areas mentioned above) are produced through a building-block approach known as the "Handbook method." This procedure also uses data from several sources, including the CPS, the CES program, State UI systems, and the decennial census, to create estimates that are adjusted to the statewide measures of employment and unemployment. Below the labor market area level, estimates are prepared using disaggregation techniques based on inputs from the decennial census, annual population estimates, and current UI data.

Attachment B: CENSUS 2000 DATA PRODUCTS

Census 2000 data products are designed to meet a variety of data needs for different segments of the data user community. The data products described here provide a summary of the general tabulation and publication program for the 50 states, the District of Columbia, and Puerto Rico (which is treated as a state equivalent for each data product). Please note that constraints with staffing and budget, federal guidelines regarding the tabulation of data by race and ethnicity, data processing, or other considerations may result in changes to the types of data products prepared or the timing of their release.

For more information on Census 2000 data products, please contact Louisa Miller (Population Division) on 301-457-2073 or by e-mail at <lmiller@census.gov>

Planned Release Date	100-Percent Data Products	Lowest Level Geography
MAR – APR 1, 2001	**Census 2000 Redistricting Data Summary File** • State population counts for legislative redistricting *Media: Internet, CD-ROM*	Blocks
JUN – SEP 2001	**Demographic Profile** • Population totals and selected population and housing characteristics in a single table *Media: Internet, CD-ROM, paper*	Places
JUN – SEP 2001	**Congressional District Demographic Profile** • Population totals and selected population and housing characteristics in a single table for Congressional Districts only *Media: Internet, CD-ROM, paper*	Congressional Districts of the 106th Congress
JUL 2001	**Race and Hispanic or Latino Summary Table CD-ROM** Medium: CD-ROM	Governmental units
At the state level: JUN – SEP 2001 *At the national level:* MAY – JUN 2002	**Summary File 1 (SF 1)**: • Population counts for 63 race categories and Hispanic or Latino….. • Population counts for many detailed race and Hispanic or Latino categories, and American Indian and Alaska Native tribes……. • Selected population and housing characteristics……. [National file includes urban and rural data] *Media: Internet, CD-ROM*	Blocks Census tracts Blocks/Census tracts
At the state level: OCT – DEC 2001 *At the national level:* JUN – JUL 2002	**Summary File 2 (SF 2)**: • Population and housing characteristics iterated for many detailed race and Hispanic or Latino categories, and American Indian and Alaska Native tribes [National file includes urban and rural data] *Media: Internet, CD-ROM*	Census tracts

At the state level: APR – DEC 2001 *At the national level:* MAY – JUL 2002	**Quick Tables** • Table shells with population and housing characteristics where the user can specify a geographic area and a population group *Medium: Internet*	Census tracts
At the state level: APR – OCT 2001 *At the national level:* JUN – JUL 2002	**Geographic Comparison Tables** • Population and housing characteristics for a list of geographic areas (e.g., all counties in a state) *Medium: Internet*	Places of 1,000 or more population
SEP – DEC 2001 (Release subject to policy decisions on access and confidentiality)	**Advanced Query Function** • User specifies contents of tabulations from full microdata file • Includes safeguards against disclosure of identifying information about individuals and housing units *Medium: Internet*	User defined down to block groups
JAN – NOV 2002	*Census 2000: Summary Population and Housing Characteristics* *Media: Internet, paper (printed report)*	Places
2003	Census 2000: Population and Housing Unit Counts *Media: Internet, paper (printed report with selected historical counts)*	Places

Continued Attachment B: Census 2000 Data Products (2)

Planned Release Date	Sample Data Products	Lowest Level Geography
DEC 2001 – MAR 2002	**Demographic Profile** • Demographic, social, economic, and housing characteristics presented in three separate tables Media: Internet, CD-ROM, paper	Places
DEC 2001 – MAR 2002	**Congressional District Demographic Profile** • Demographic, social, economic, and housing characteristics presented in three separate tables for Congressional Districts only *Media: Internet, CD-ROM, paper*	Congressional Districts of the 106th Congress
AUG – DEC 2002	**Summary File 3 (SF 3)**: • Population counts for ancestry groups…… • Selected population and housing characteristics……..….. *Media: Internet, CD-ROM*	Census tracts Block groups/ Census tracts
DEC 2002 – MAR 2003	**Summary File 4 (SF 4)**: • Population and housing characteristics iterated for many detailed race and Hispanic or Latino categories, American Indian and Alaska Native tribes, and ancestry groups Media: Internet, CD-ROM	Census tracts
AUG 2002 – MAR 2003	**Quick Tables** • Table shells with population and housing characteristics where the user can specify a geographic area and a population group Medium: Internet	Census tracts
SEP 2002 – JAN 2003	**Geographic Comparison Tables** • Population and housing characteristics for a list of geographic areas (e.g., all counties in a state) *Medium: Internet*	Places of 1,000 or more population
For 5-percent sample: APR – JUL 2002 For 1-percent sample: AUG – DEC 2002	**Public Use Microdata Sample (PUMS) Files** • 5-percent sample (information for state and sub-state areas) • 1-percent sample (information for metropolitan areas) *Medium: CD-ROM*	Public use microdata areas (PUMAs)
DEC 2002 – MAR 2003 (Release subject to policy decisions on access and confidentiality)	**Advanced Query Function** • User specifies contents of tabulations from full microdata file • Includes safeguards against disclosure of identifying information about individuals and housing units Medium: Internet	User defined down to census tracts
2003	*Census 2000: Summary Social, Economic, and Housing Characteristics* Media: Internet, paper (printed report)	Places
2003	**Congressional District Data Summary File** • 100-percent and sample data for the redistricted 108th Congress Media: Internet, CD-ROM	Census tracts within Congress-ional Districts

(Revised: 2/1/00)

Attachment C

Examples of Relevant Web Sites

AMERICAN COMMUNITY SURVEY: **www.census.gov/acs/www/**
Summarized data updated each year for areas; documentation for the survey, including a questionnaire

ANNIE E. CASEY FOUNDATION: www.ecf.org
The Foundation's mission is to help build better futures for millions of disadvantaged children who are at risk of poor educational, economic, social, and health outcomes.

BUREAU OF THE CENSUS: www.census.gov
Click on "American FactFinder" for data from censuses and surveys.

BUREAU OF LABOR STATISTICS, LOCAL AREA UNEMPLOYMENT STATISTICS:
www.bls.gov/lauov.htm

CENTER ON URBAN POVERTY AND SOCIAL CHANGE, Case Western Reserve University (contact: Claudia Coulton): **http://povertycenter.cwru.edu**
Cleveland Area Network for Data Organizing, a data base that contains neighborhood-level information from the 1990 census and a variety of administrative data files.

CENTERS FOR DISEASE CONTROL AND PREVENTION:
http://www.cdc.gov/scientific.htm

CHILD TRENDS: www.childtrends.org

CLEVELAND AREA NETWORK FOR DATA AND ORGANIZING (CAN DO):
http://povertycenter.cwru.edu
Center on Urban Poverty and Social Change, Case Western Reserve University. This is an interactive database of Cleveland neighborhoods with mapping capabilities

COUNCIL OF PROFESSIONAL ASSOCIATIONS ON FEDERAL STATISTICS
http://members.aol.com/copafs

DC AGENDA: **http://www.dcagenda.org**
Neighborhood Information Service assembles neighborhood-level data and analyses to support community-based organizations.

DEMOGRAPHIC DATA VIEWER
http://plue.sedac.ciesin.org/plue/ddviewer/ddv20/htmls/map-intro.html

DEPARTMENT OF AGRICULTURE, Economic Research Service
http://www.ers.usda.gov/rural

Rural briefing room.

DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT, Office of Policy Development and Research:
http://www.huduser.org/datasets/pdrdatas.html
Includes links to varous data sets associated with housing and urban development such as the American Housing Survey and Low Income Housing. Includes information such as annual adjustment factors to project-based rent subsidies, assisted housing (national and aggregated information on housing assistance at the local level, fair market rents, government sponsored enterprise data, income limits, low-income housing tax credit, property owners and managers survey, qualified census tracts and difficult development areas, and section 8 administrative fees.

HUD's State of the Cities Data System: **http://webprod.aspensys.com/socds/**
Contains data sets for tracking the conditions of America's cities.

DEPARTMENT OF JUSTICE
NATIONAL INSTITUTE OF JUSTICE: **www.ojp.usdoj.gov/nij**
CRIME MAPPING RESEARCH CENTER: **www.ojp.usdoj.gov/cmrc**

ECONOMIC DEVELOPMENT ADMINISTRATION: www.doc.gov/eda
Obtain their report on "Socioeconomic Data for Economic Development: An Assessment" by Andrew Reamer and Joseph Cortright at: **www.doc.gov/eda/pdf/socio.pdf**

ESRI for GIS mapping and software: **www.esri.com**

FEDSTATS: www.fedstats.gov
Nationally available statistics produced by the federal government.

FLORIDA DEPARTMENT OF CHILDREN AND FAMILIES:
http://www.state.fl.us/cf_web/district11
Their GIS Research Services unit focuses on data collection, analysis, neighborhood indicators and community mapping.

HEALTHY COMMUNITIES PROJECT, UNVERSITY OF TEXAS-HOUSTON, SCHOOL OF PUBLIC HEALTH: www.sph.uth.tmc.edu/chis/background.htm

INTEGRATED DATABASE ON CHILDREN'S SERVICES IN ILLINOIS, CHAPIN HALL CENTER FOR CHILDREN, UNIVERSITY OF CHICAGO:
www.chapin.uchicago.edu
How children in Illinois use public social service programs. This data base includes some longitudinal data and is being used as a national standard for how to use administrative records for policy. Replication is planned in California, Massachusetts, North Carolina, and Wisconsin.

MARYLAND GOVERNOR'S OFFICE FOR CHILDREN, YOUTH, AND FAMILIES
www.ocyf.state.md.us/results.htm

Tracks results and indicators of the well being of Maryland's children.

NATIONAL NEIGHBORHOOD INDICATORS PROJECT, URBAN INSTITUTE, WASHINGTON, D.C.
http://www.urban.org/nnip/index.htm
The National Neighborhood Indicators Partnership (NNIP) is a collaborative effort by the Urban Institute and local partners to further the development and use of neighborhood-level information systems in local policymaking and community building. All local partners have built locally self-sustaining information systems with integrated and recurrently updated information on Neighborhood conditions in their cities. These systems facilitate the direct use of information by local government and community leaders to build the capacities of distressed urban neighborhoods.

PITON FOUNDATION, DENVER: www.piton.org
This is a private foundation whose mission is to provide opportunities for children and their families to move from poverty and dependence to self reliance. The Piton's Data Initiative for Denver has the user friendly Neighborhood Facts, a web-based database of neighborhood indicators, maps, and histories.

POPULATION ESTIMATES, Bureau of the Census:
http://www.census.gov/population/www/estimates/

RAY MARSHALL CENTER FOR THE STUDY OF HUMAN RESOURCES (LBJ School of Public Affairs, U. of Texas): www.utexas.edu/research/cshr/

ROCKEFELLER FOUNDATION: www.rockfound.org

SAN DIEGO ASSOCIATION OF GOVERNMENTS:
http://www.sandag.cog.ca.us/data_services/
SANDAG maintains extensive databases of facts and figures from a historical perspective, current information, and forecasted information. The information encompasses population growth, housing, employment, and income as well as data on crime and the local economy.

SMALL AREA INCOME AND POVERTY ESTIMATES (SAIPE):
www.census.gov/hhes/www/saipe/techdoc
Detailed information about the SAIPE program is available on the Census Bureau's web page.

UNIVERSITY OF MICHIGAN, Inter-University Consortium for Political and Social Research (ICPSR): < **http://www.icpsr.umich.edu/archive1.html**>
Archive of computerized social science data.

UNIVERSITY OF WISCONSIN AT MADISON, INSTITUTE FOR RESEARCH ON POVERTY: www.ssc.wisc.edu/irp/
The Institute for Research on Poverty is a national, university-based center for research into the causes and consequences of poverty and social inequality in the United States. It is nonprofit and nonpartisan.

URBAN INSTITUTE: **www.urbaninstitute.org**
URBAN INSTITUTE'S NATIONAL NEIGHBORHOOD INDICATORS PARTNERSHIP: www.urbaninstitute.org/nnip/
The NNIP is a collaborative effort by the Urban Institute and local partners to develop and use neighborhood-level information systems in local policymaking and community building. The local partners have built locally self-sustaining information systems with integrated and recurrently updated information on neighborhood conditions in their cities.

URBAN STRATEGIES COUNCIL, Oakland California: **http://www.urbanstrategies.org**
The data group used administrative records files and census files for city reports and maps.