

Catalog of Administrative Data Sources :

FOR NEIGHBORHOOD INDICATOR SYSTEMS

By: Claudia J. Coulton, with Lisa Nelson and
Peter Tatian, September 1997¹

National Neighborhood Indicators Partnership

THE URBAN INSTITUTE

The nonpartisan Urban Institute publishes studies, reports, and books on timely topics worthy of public consideration. The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders.

CONTENTS

| | |
|--|-----------|
| INTRODUCTION | 1 |
| ECONOMY | 1 |
| Employment Services | 1 |
| Business Directories | 9 |
| Municipal Income Tax Records | 10 |
| Training | 10 |
| EDUCATION | 12 |
| Public School Records | 12 |
| Head Start Records | 13 |
| HEALTH | 13 |
| Vital Records | 13 |
| Communicable Diseases | 15 |
| Emergency Medical Services | 15 |
| Immunization | 16 |
| Medicaid Claims | 16 |
| Hospital Discharge Files | 17 |
| SOCIAL SERVICES | 17 |
| Public Assistance Files | 18 |
| Subsidized Day Care | 18 |
| Day Care Licenses | 19 |
| Child Welfare Records | 19 |
| Mental Health, Alcohol, and Drug Services | 20 |
| SAFETY AND SECURITY | 20 |
| Municipal Police | 21 |
| Emergency 911 Call Data | 22 |
| Juvenile Court | 23 |
| Coroners' Reports | 23 |
| Child Maltreatment | 24 |
| Liquor Licenses | 25 |
| COMMUNITY RESOURCES AND PARTICIPATION | 25 |
| Voter Records | 25 |
| Membership Records | 26 |
| Public Transit | 26 |
| Auto Registration and Licenses | 26 |
| Community Development Block Grants | 28 |
| Community Directories | 29 |

| | |
|--|-----------|
| HOUSING | 29 |
| Home Mortgage Disclosure Act (HMDA) | 29 |
| Property Data from Local Tax Assessor or Auditor | 31 |
| Building and Demolition Permits | 32 |
| Housing Code Enforcement | 32 |
| Public and Subsidized Housing | 33 |
| | |
| THE ENVIRONMENT | 34 |
| Air Pollution | 35 |
| Water-Discharge | 36 |
| Superfund Sites | 37 |
| Toxic Chemicals | 38 |
| Hazardous Waste | 38 |
| | |
| ISSUES IN USING ADMINISTRATIVE DATA | 39 |
| Geographic Boundaries | 39 |
| Confidentiality | 39 |
| Data Accuracy | 40 |
| Data Extraction and Management | 41 |
| Matched and Longitudinal Files | 42 |
| | |
| ENDNOTES | 44 |
| | |
| REFERENCES | 45 |

INTRODUCTION

There is a long tradition of using data collected for administrative purposes to produce social and economic indicators (Rossi, 1972; Annie E. Casey Foundation, 1997). The outcome-based accountability thrust of many social programs has raised additional interest in such information (Schorr, 1994). Many administrative agencies now have computerized records and information systems. Further, the advent of geographic information system (GIS) technology now makes it feasible to calculate many indicators for small areas such as communities and neighborhoods from these large databases.

There are many sources and types of data from administrative agencies that can be used to produce measures useful to community-based organizations. This monograph provides a description of these data sources because such information is not readily available in other places. Most databases are maintained by local agencies, but there are a few state and federal databases that can be used for small area measures. Because of the local nature of much of the data, the descriptions provided here may not match exactly what is to be found in a particular locale.

Although the list is long, it is not exhaustive. The sources are grouped into seven categories: economy, education, health, social services, safety and security, community resources and involvement, housing and physical development, and the environment. This reflects the primary use to which the data have been put. However, many of these data sources, when combined with other information, can prove valuable across categories. (See summary in Table 1.)

After discussing individual data sources in these categories, we close with a brief section discussing issues in the use of administrative data sources like these. This section tells users what they should watch out for when using such files and offers guidance to avoid mishaps.

ECONOMY

Many community organizations have economic development as an aspect of their work. Information about the economic and job activity in an area can be obtained through a variety of administrative data sources.

Employment Services

Departments of employment services are the primary sources of employment information in states. These departments, though, do not usually publish information for neighborhoods. Nevertheless, when subjected to geographic analyses, several of their databases may be useful in measuring aspects of employment within neighborhoods.

ES202: Each state is federally mandated to collect reports related to the unemployment insurance program from every establishment that employs more than one worker. From these reports, the *ES202* database is generated. This database covers nearly all employers with paid employees. Self-employed individuals such as doctors and attorneys and family businesses with no paid employees may be missing from this database. This is the only government database that provides company-level information along with a geographic location.

TABLE 1
SUMMARY OF ADMINISTRATIVE DATA SOURCES FOR SMALL AREA MEASURES

| Data Source | Indicator Examples | Availability |
|---|--|--|
| ECONOMY | | |
| ES202: Employers' quarterly reports of industry classification, total employment, and payroll | % change in total employment % employment in manufacturing | These are confidential employer reports filed with the state Employment Services Agency. In many states they can be obtained for legitimate research purposes with proper safeguards. |
| UI wage record: Employee earnings and weeks worked | % former welfare recipients who found jobs Average earnings of participants in training | These are confidential individual records filed by employers with the state Employment Services Agency. There are no addresses on file, so they must be merged with another data source for small area analysis. Confidentiality protection as in ES202 above. |
| UI Claimant File: Applicants for unemployment benefits | # of UI claims filed Average weeks of benefits paid | These are confidential individual records of individual claims filled with the state Employment Services Agency. |
| Business Directories: Selected listings of businesses such as Cole's, Harris, Dun and Bradstreet | # of establishments by industry % increase in establishments | Commercially available from vendors. |
| Municipal Income Tax Records: Employers' withholding records | Total dollars generated by employers Increase in tax dollars | These are highly confidential data but municipal taxing agencies may respond to requests for information aggregated to the neighborhood level. |
| Public Training: Records of trainees in public programs | # of persons in training % of trainees placed in employment | These are confidential records of trainees. Records for JOBS are maintained by the state or local Human Service Departments. JTPA records are maintained by the state Employment Services Agency. |
| EDUCATION | | |
| Public school records: Personnel records of students enrolled in public schools | % children absent more than 20% of the days % children passing 9 th grade proficiency exam | Confidential records maintained by Board of Education. Availability and format vary by jurisdiction. Information reported to the state is usually aggregated to school building or school district and is typically unsuited to small area analysis. |
| School entry records: Individual records at enrollment | % entering kindergarten with pre-school | Confidential records maintained by Board of Education. Availability and format vary by jurisdiction. |
| Head Start records: Individual records of students enrolled in Head Start | % eligible children attending Head Start Average number of months enrolled | Confidential records maintained by local Head Start Agency. There may be more than one such agency in a city. Availability varies. |

TABLE 1
SUMMARY OF ADMINISTRATIVE DATA SOURCES FOR SMALL AREA MEASURES

| Data Source | Indicator Examples | Availability |
|---|---|--|
| HEALTH | | |
| Vital records: Birth, death, and fetal death certificates | % pregnancies with adequate pre-natal care Infant mortality rate Teen birth rate % births to unmarried mothers | Available annually from state vital statistics office. May already have census geographic codes. Index portion of certificate with address available under confidentiality agreement for valid research purposes. Matched birth-death file available but often delayed by several years. |
| Communicable diseases: Records of reportable disease incidence | # of cases of sexually transmitted diseases | Confidential records maintained by state and local health officials. Availability and format differ by jurisdiction. |
| School immunization records: Records of student immunization status | % kindergartners appropriately immunized by age 2 | Confidential records maintained by local school boards. Availability and format vary by jurisdiction. |
| Medicaid claims: Claims submitted for medical services | Annualized rates of ambulatory care utilization % care obtained from emergency rooms | Confidential records maintained by state agency that administers Medicaid. Availability and format vary by state. As states move to managed care claims, these records may no longer be generated. Managed care reporting requirements may not be geographically specific. |
| Hospital discharge files: Reports on hospital discharges | Age-adjusted rates of hospitalization | Confidential information published at zip code level by some hospital associations. Availability varies. Rarity of hospital discharge may make rates unreliable for small areas. |
| SOCIAL SERVICES | | |
| Public assistance records: Monthly eligibility and payment files | # of recipients per 100 population Total dollars of benefits paid per month % of long-term recipients | Confidential records maintained by local or state Human Services Agency. Format and availability vary. Multiple months must be linked to determine duration of welfare reciprocity. |
| Subsidized day care: Records of children receiving day care subsidies | # of children in subsidized day care % vouchers used in day care homes | Confidential records maintained by local or state Human Services Agency. Format and availability vary. |
| Day care licensees: Records of licensed day care providers | # of day care slots by type % increase in day care slots | Confidential records maintained by state licensing authority or local child care resource and referral agencies. Availability and format vary. |
| Child welfare records: Records of children in custody or receiving protective services | # of children taken into care annually # of children under agency supervision # of children in foster placement | Confidential records maintained by local and state child welfare agency. Availability and format vary. |

TABLE 1
SUMMARY OF ADMINISTRATIVE DATA SOURCES FOR SMALL AREA MEASURES

| Data Source | Indicator Examples | Availability |
|---|---|---|
| Mental health, alcohol, and other drugs services records: Records of individuals receiving services through local mental health and substance abuse treatment systems | # of persons in treatment % of clients keeping regular visits | Confidential records maintained by local boards. Availability and format vary. |
| SAFETY AND SECURITY | | |
| Police records: Records of crime reports, victims, suspects, and arrests | Crime rates % crimes committed by residents versus outsiders | Maintained by municipal police departments. Availability varies by jurisdiction. Most departments follow Uniform Crime Reporting procedures. |
| Juvenile court filings: Records of charges in court | Juvenile crime rate % juvenile filings that are for violent offense | Confidential records maintained by county court. Availability and format vary by jurisdiction. |
| 9-1-1 calls: Calls to the local emergency dispatch system | # of calls for domestic violence % change in calls | Maintained by regional agency. Availability varies by jurisdiction. |
| Coroner's reports: Records of deaths | % homicides involving firearms # of suicides involving drugs | Maintained by local coroner. Availability varies by jurisdiction. May not be computerized. |
| Child abuse and neglect reports: Records of reports to child protection authorities | % children with maltreatment reports % reports that involve neglect | Confidential records maintained by local or state child protection agencies. Availability limited and requires strict confidentiality protection. |
| Liquor licenses: Permits to sell alcoholic beverages | # and type of outlets | Public record maintained by the state Bureau of Liquor Control. |
| COMMUNITY RESOURCES AND PARTICIPATION | | |
| Voter records: Voter registration | % of eligible voters who are registered | Records maintained by local boards of elections. Available for small areas. |
| Membership records: Local organizations' membership records | % neighborhood residents who are members | Organizations vary greatly in record keeping and willingness to release information. If addresses are recorded, small area measures can be developed. |
| Community directories: Libraries, churches, parks, playgrounds, recreation centers, and other listings | # of institutions | Directories vary in accuracy and specificity. Computerized yellow pages has wide coverage but includes only those organizations that choose to be listed there. |
| Community Development Block Grants: Records of funding expenditures | dollars expended annually % expenditures by category | Available from local CDBG administrator. Difficult to allocate to small areas because many of the expenditures are not neighborhood specific. |
| Public transit records: Routes and schedules and ridership | Average travel time to key locations Number of direct routes to neighborhood | Available from local transit authority. Travel times and distances will often need to be calculated using mapping software. |
| Automobile registrations and licensees: Records for vehicles and drivers | # of automobiles per capita % of population with valid drivers licenses | Available for a fee from the State Bureau of Motor Vehicles. |
| HOUSING | | |

TABLE 1
SUMMARY OF ADMINISTRATIVE DATA SOURCES FOR SMALL AREA MEASURES

| Data Source | Indicator Examples | Availability |
|---|--|---|
| HMDA: Records of loan applications and approvals | % conventional loans % loans denied # of financial institutions making loans | This is public information. Available annually from the Federal Reserve Bank's Financial Institutions Council |
| Local tax assessor: Real property records maintained for taxing purposes | % properties tax delinquent Median housing assessed values % parcels that are residential | These are public records. Available from local tax assessor. Format and ease of use vary by jurisdiction. |
| Building and demolition permits: Records of permits issued | Total value of construction # of buildings demolished | These are public records. Available from city building departments. May not be computerized. |
| Housing code enforcement: Records of violations | % properties with violations | These are public records. Available from city building inspection department. May not be computerized. |
| Public and subsidized housing: Records of units of housing | # of public housing units persons/unit in public housing | The data are available annually from HUD in computerized format. |
| ENVIRONMENT | | |
| Air pollution facilities records: Emissions and compliance data for air pollution point sources | Location of pollution sources Amount of emissions from sources Location of sources not in compliance with emission standards | Available from the Environmental Protection Agency (EPA) through the Envirofacts website, AIRS Executive software, or directly from the Office of Air Quality Planning and Standards. |
| Air quality data: Levels of pollution observed at monitoring stations throughout the U.S. | Amount of air pollutants at different monitoring locations | Available through AIRS Executive software or directly from EPA's Office of Air Quality Planning and Standards. |
| Water-discharge: Records on water-discharge permits regulated under the Clean Water Act | Location of wastewater discharge sources Location of sources in violation of discharge regulations | Available from EPA through the Envirofacts website. |
| Superfund sites: Records on Superfund cleanup sites | Location of Superfund sites Type of incident for site Current enforcement status Current activity status | Available from EPA through the Envirofacts website or through Superfund website. |
| Toxic chemicals: Data about the release and transfer of toxic chemicals | Locations of toxic chemical releases Amount of release Type of chemical released Medium of release (air, water, etc.) | Available from EPA through the Envirofacts website or through Superfund website. |
| Hazardous waste: Data about generators, transporters, treaters, storers, and disposers of hazardous waste | Locations of toxic waste facilities Locations where corrective action has been taken Locations of openings and closings of toxic waste sites | Available from EPA through the Envirofacts website or through Superfund website. |
| Notes: UI=unemployment insurance; JOBS=Job Opportunities and Basic Skills; JTPA=Job Training Partnership Act; CDBG=Community Development Block Grant; HMDA=Home Mortgage Disclosure Act; HUD=U.S. Department of Housing and Urban Development; AIRS=Aerometric Information Retrieval System | | |

The data contained in the ES202 file include establishment name, legal name, address, city, zip, state, county code, four-digit industry (standard industrial classification) code, ownership code (indicating public or private ownership), number of employees, and total quarterly wages. Total monthly employment is provided for each establishment. There are two variables—unemployment insurance (UI) number and reporting unit number (RUN)—that uniquely identify a company. There is a date that indicates when the UI number was established. There are a successor UI and a predecessor UI that enable the user to link companies over time that may have changed ownership. However, these variables are not specifically required and therefore may not be reported. Additionally, there is a variable that indicates whether the company is a multi-unit employer (e.g., a bank) or a single-unit employer. The data are available quarterly from the state's Bureau of Employment Services. The first quarter of data are available in the same quarter of the following year. The data are available electronically; however, the format may vary by state.

While it provides unique establishment level information, there are some well-known problems with ES202 data (White, Zipp and McMahon, 1990; Leete and Bania, 1995; Waits, Rex and Melnick, 1997). One problem is the accuracy of the establishment's address information. Even though firms are instructed to use the address of the actual location, some firms mistakenly report all employment at the address of a headquarters or list the address of the accountant who completes the report. There are additional problems if the user wants to look at establishments over time. The establishment has the unique UI and RUN numbers to link quarter by quarter; however, if an establishment changes ownership, these numbers will change. If the predecessor UI and/or successor UI are not reported, there may be some difficulty linking establishments.

The advantage of the ES202 data is the accuracy of the total employment that is reported by each firm. The federal mandate carries with it a tax liability that encourages accurate reporting of employment. In addition, the ES202 employment counts at the county level have been found to closely track other official government measures of employment.

There are federal and state restrictions on who may obtain this data and how it may be used. It should also be noted that there are some confidentiality rules that must be adhered to when reporting the ES202 data. If one establishment in any given standard industrial classification (SIC) code and any given geographic area accounts for more than 80% of the employment, the data must be suppressed. In addition, if an establishment is one of only two establishments within a given SIC code and given geographic area, the data must be suppressed.

The ES202 data can be used to calculate several community measures of employment. It should be noted that these measures pertain to people working in neighborhood firms, not to neighborhood residents. Total employees, wages paid, and number of establishments can be calculated by SIC codes and various geographic areas. Average payroll per employee by industry and geographic area can be calculated. If data are analyzed over time, the number of business openings, closings, relocations, and associated changes in employment and wages can be

determined. Employment gains and losses by geographic area and SIC codes can also be computed. The ES202 employment data have been used to measure employment change by SIC in small geographic areas, develop estimates of employment in various geographic areas, determine location of specific types of employment, and estimate location of expected job openings.

UI wage record: The UI wage record is also available from a state's Bureau of Employment Services. It is collected under the same federal mandate as ES202. In addition to information about the employer, the wage record file contains specific information about the employee.

The wage record contains employee information such as name, social security number, quarterly wages paid, and weeks worked. Additionally, there is information about the employer such as name, address, city, zip code, state, industry code (SIC), and the employer's UI number (described under ES202). These data are also available quarterly. Like the ES202 data, there are federal and state restrictions on obtaining and using the wage record file.

The UI wage records cannot be used alone to develop neighborhood indicators because they do not contain individual wage earners' home addresses. However, public assistance records that are address coded have been linked to UI records to determine employment experiences of welfare recipients within a geographic area. Links could be made to other agency records as well.

UI claimant file: The UI claimant file is collected under the same federal mandate as ES202 and the UI wage record. This file contains specific information about the individuals who have filed for unemployment compensation.

The data available from this file include the claimant's social security number, address, birthdate, sex, and race. Additional information includes weekly benefit amount, average weekly wage, number of qualifying weeks, date of claim, date of separation from job, pay rate, and employer UI account number. These data are confidential and state and federal regulations restrict who may obtain these data and how they may be used. The data are available quarterly from the state's Bureau of Employment Services.

The community measures that can be developed from these data include percentage of persons receiving benefits, average length of time on unemployment, and percentage who have exhausted benefits. Demographic characteristics of unemployment compensation recipients can also be computed. The data have been used, in conjunction with other data sources, to target recipients who have exhausted benefits and to provide access to job training programs.

Business Directories

There are many business directories that provide employment information. The information available varies by directory. Most libraries will have a list of the directories available and the types

of information each directory provides. These directories may not be a complete source of employment information; however, used in conjunction with ES202, they may enhance the accuracy of addresses and establishments in a given geographic area (Carlson, 1995). A few of the business directories are discussed below.

Cole's Business Directories: This directory of businesses is published for each state and distributed by Cole Publications. It is available in printed form and electronic form (CD-ROM) and is published by the American Directory Publishing Company in Omaha, Nebraska. There are several drawbacks to this database. Unlike ES202, it contains only employment ranges, not actual employment levels, and firms are under no legal requirement to be listed in the directory. In addition, while an attempt is made to update and verify every listing, this occurs continuously, making it difficult to get a listing of firms at a single point in time. The listing is based on telephone directories, so firms not listed in the telephone book are unlikely to be included in this database.

Various versions of this database are sold as electronic "national yellow page" listings for all businesses. This directory is aimed at businesses involved in marketing products to firms in particular industrial classes. There are other electronic directories of businesses, including ProPhone Business Listing, PhoneDisc PowerFinder, and Select Phone Business Listings. These directories have the same problems as *Cole's Business Directories*.

Dun and Bradstreet: The Dun Market Indicators (DMI) file is based on credit-rating data of Dun and Bradstreet. Therefore, the file is limited to those firms that actively seek out a credit-rating record with Dun and Bradstreet. Smaller firms and those without a credit history are likely to be excluded. DMI will be much more likely to include outdated data, since firms that close or move have no incentive to update their records. Additionally, firms may understate or overstate employment levels in an effort to improve credit ratings. DMI does not attempt to track employment for every establishment location; therefore, employment counts may represent county or area totals for a given firm.

Harris Directory: The *Harris Directory* is published annually by Harris Publishing Company of Twinsburg, Ohio. This file is available only for selected midwestern and southern states. In addition, this directory contains mainly manufacturing companies. As with DMI, employment is not necessarily reported by establishment location; therefore, the employment count may represent a firm total. It may also contain some of the same problems as the other business directories discussed.

Municipal Income Tax Records

Many states allow the municipalities or counties to collect taxes on the income earned by the residents living in the municipality or county and nonresidents working in the municipality or county.

These taxes are collected by the municipality imposing the tax or by a central collection agency on behalf of the municipality.

The municipal taxes are withheld by the employer and provided to the appropriate tax collector. The information collected by the municipality or collection agency includes the employer name and address, total amount of local taxes withheld from employees, taxes paid based on net profit of business, and a federal identification number. Generally, the taxes must be paid on a monthly or quarterly basis to the municipality or collection agency. Because of the confidentiality of these data, the ability to obtain this information will vary by municipality or collection agency. Typically, aggregate data would need to be produced by the tax collection agency for selected neighborhoods in response to a special request.

The small area indicators that can be developed from these data include income tax generated by businesses in a given geographic area. If confidentiality agreements can be crafted to cover it, it might be possible to link these data with other data sources such as ES202 using the federal identification number. This would allow additional indicators to be developed, such as income tax generated by industry code and total wages and number of employees by geographic area and industry code.

Training

Federally funded training programs generate data that can be used to determine the levels of participation in these activities in local communities. These programs are likely to undergo significant change in the next several years because of block grants and welfare reform. Nevertheless, the training databases are likely to continue with similar information.

Job Opportunities and Basic Skills: Job Opportunities and Basic Skills (JOBS) has been the federally mandated program aimed at assisting families in making the transition from welfare to self-support. The JOBS activities include job search, work experience, education, and training services.

The JOBS program was governed by federal and state regulations and, in most cases, the data system has been operated by the department responsible for public assistance administration. Welfare-to-work programs will continue under welfare reform, and states and localities are likely to maintain similar, perhaps improved, databases.

The JOBS information is contained in two files. One file contains demographic information about the public-assistance recipient obtaining JOBS services and the other file contains information about the recipient's activity history in the JOBS program. The demographic information includes social security number; address including street, city, zip code, and county; race; sex; date of birth; educational level; whether the recipients attend school; marital status; cash benefit from Aid to Families with Dependent Children (AFDC); and length of time on assistance. The JOBS program information includes a chronological history of an individual's participation status including

eligibility/ineligibility, job assessment, job assignment, attendance record, exemptions (i.e., long-term illness, very young children), failure to participate, sanctions, and employment history. These two files can be linked by social security number. These data contain confidential information about each individual participating in the program. The ability to obtain the information may vary depending on the purpose for which it is sought. Monthly participation data are available.

The small area indicators that can be developed from the JOBS data include the average time between a job assessment and the assignment to an activity, the participation rate among recipients, participation rates by JOBS activity, sanction rates, and employment rates of participants. The data from the JOBS program have been used to determine participation rates in welfare-to-work programs. In addition, the JOBS data have been used in conjunction with public assistance records to determine whether long-term recipients have participated in JOBS activities (Coulton, Verma and Guo, 1996).

Job Training Partnership Act: The Job Training Partnership Act of 1982 (JTPA) established the nation's largest employment and training program for disadvantaged adults and youth. The purpose of this act is to establish programs that prepare adults and youth facing serious barriers to employment. Program participation is voluntary. Potential participants must meet certain criteria to receive JTPA services. The JTPA program provides classroom vocational training, on-the-job training, job search assistance, and other related training services. The goals of this program are to increase employment and earnings and reduce welfare dependence. This program is administered by the state, and service delivery is provided locally. The data available from the JTPA program include information about the participant such as social security number, birthdate, sex, race, address, eligibility status, employment history and status, educational status, and participation in public assistance programs. In addition, information regarding application date, termination date, and services is provided. There is also information regarding the aptitude, ability, and skill levels of program participants. Employment and public assistance status is also provided approximately three months following termination from the program. The data are collected and maintained by the state's Bureau of Employment Services. The data are available quarterly. Like the other employment-related data files, there are restrictions on who may obtain these data and how they may be used.

The small area indicators that can be developed from these data include participation rates, participation rates by race and sex, percentage of participants on welfare, and the educational status of program participants. Additionally, it is possible to calculate the percentage who drop out of the program, the skill levels of participants, educational status of participants, percentage of participants who become employed, average wage when employed, and percentage of participants whose welfare dependency was reduced. The data have been used to document who is participating and outcomes of program participation such as whether welfare dependency has decreased and whether employment has occurred.

EDUCATION

Educational outcomes are important indicators of the well-being of the community and the functioning of its systems. The focus has traditionally been on children and youth because education is a fundamental component of development (Thornton, Love and Meckstroth, 1994). School systems typically generate overall reports for the state and community but seldom produce data at the neighborhood level.

Public School Records

Most public school districts maintain computerized files of individual student records. These records are confidential but, with proper protection agreements, can be used to develop measures for small areas. These files include the student's address, the school attended, school transfers or leavings, scores on standardized achievement and proficiency tests, attendance and disciplinary records, free lunch eligibility, and family status. Data availability is limited for private and parochial schools because most of them maintain separate and unique record systems.

The data can be used to calculate attendance rates and average achievement for students by school or by neighborhood. School and residential mobility can be calculated by matching students' records across years to determine movement. School completion requires matching records for a cohort of students, usually from the eighth grade onward, to determine those who graduate. School entry records have also been used to determine immunization status and school readiness. The nature of these records varies considerably across districts.

Head Start Records

Area Head Start agencies maintain records on children enrolled in Head Start and on the individual Head Start programs. These files contain the child's address, the Center location, enrollment dates, and other selected family and child information. These records are confidential and their format and availability differ across local agencies.

Head Start records can be geo-coded and aggregated to calculate rates of Head Start enrollment in small areas and distances between home and Centers for residents. Matched with school enrollment files, they can be used to calculate Head Start participation rates among children entering school.

HEALTH

Health is defined as not merely the absence of disease but the overall physical, mental, and social well-being of a person. Vast improvements have been made in the area of preventive health in the last 50 years, especially in developing nations, but the indicators for many low-income communities compare unfavorably with the overall high standard of health in the United States (Geronimus, Bound, Waidmann, Hillemeier and Burns, 1996). There is a tradition of examining health using small area data (Gould, Mahajan and Lucero, 1989).

Vital Records

The registration of births, deaths, fetal deaths, and other vital events is a state and local function. The civil laws of every state provide for a continuous, permanent, and compulsory vital registration system.

The state vital statistics office issues certificates of live birth, fetal death, or death either directly or through a local registrar. A live birth is defined as the complete expulsion or extraction from its mother of a product of human conception that, after such expulsion or extraction, breathes or shows any other evidence of life such as beating of the heart, pulsation of the umbilical cord, or definite movement of voluntary muscles, whether or not the umbilical cord has been cut or the placenta is attached. A death is defined as the permanent cessation of all vital functions. A fetal death is defined as death prior to the complete expulsion or extraction from its mother of a product of conception of at least 20 weeks gestation, which, after such expulsion or extraction, does not breathe or show any other evidence of life such as beating of the heart, pulsation of the umbilical cord, or definite movement of voluntary muscles.

The records are typically compiled by the state vital statistics office. Birth information is available in two sections. In the first section, each record contains a unique birth certificate number,

mother's name, address, and other such demographic and person-identifying information. Some localities include a census tract designation. This is called the index portion and is confidential and released only when a special request justifying the need for such information is made. The other portion is called the statistical file and has, in addition to the birth certificate number, information about prenatal care, congenital anomalies, and birth weight. Some jurisdictions also include the census tract on the statistical portion, and that locational identifier is widely available for public health research.

The death file consists of a unique death certificate number, name, social security number, and many other indicators, such as cause of death. The fetal death file has almost the same information as the birth file but also includes the cause of death.

There are many small area indicators that can be calculated from birth certificate data. Recorded birth weights can be analyzed to arrive at the number of low-birth-weight infants. Information about mother's prenatal care visits are also recorded. Kessner's adequacy of prenatal care index can be calculated from these data (Kessner, Singer, Kalk and Schlesinger, 1973). Kessner's index is arrived at by taking into account the number of gestation weeks, the timing of entry into care, and the number of total prenatal care visits. A certain number of prenatal visits are expected depending upon the gestation weeks of the infant. If the actual number of prenatal care visits matched or exceeded the expected number of visits and the mother had entered early prenatal care (in the first trimester) then the prenatal care is considered adequate. To calculate this, divide the actual number of visits by the expected number of visits to find the ratio of actual visits to expected visits. If the ratio equals or exceeds 1, the prenatal care is considered adequate.

The death file contains such information as cause of death. These can be analyzed for small areas to see if the leading causes are different from those of the nation as a whole. Infant death rates can be calculated and compared with city, state, and national figures to identify areas of comparatively high infant death rates. Excess mortality can be calculated by comparing age-specific deaths in the neighborhood with the expected deaths based on a standard population (McCord and Freeman, 1990). Prenatal information can be analyzed from the fetal deaths file to identify small geographic areas of risk for such outcomes.

Most state and local governments publish vital statistics reports and some of them contain small area data. The Department of Vital Statistics of the state and local governments can be contacted for these reports. The National Center for Health Statistics (NCHS) publishes monthly and annual reports for the nation, states, counties, cities, and regions. Selected NCHS publications can be viewed at the following website: <http://www.cdc.gov/nchswww/nchshome.htm>.

Communicable Diseases

Diseases caused by the direct or indirect spread of pathogens from one person to another are called communicable diseases. Diseases such as tuberculosis, syphilis, and AIDS are examples of communicable diseases. Many communicable diseases are reportable, under law, to local and state health officials. Physicians are responsible for filing these reports.

Although not always computerized, these data are maintained by departments of health and usually contain addresses that would allow small area analysis. However, confidentiality and generalizability issues require special justifications for release of such data. Furthermore, small geographic areas may have incidences so small that no meaningful analysis can be carried out. Examples of the types of indicators that could be developed are the incidence in a given area and the classification within the specific group of disease. Some information about the emerging and other communicable diseases can be found at the World Health Organization website at http://www.who.ch/programmes/emc/emc_home.htm. The Center for Disease Control publishes information on communicable diseases in selected metropolitan areas at <http://www.cdc.gov/publications.htm>.

Emergency Medical Services

Emergency medical services (EMS) are services that are delivered with a sense of urgency to those in need of immediate attention. Accident victims and patients having heart attacks are examples of those needing immediate medical attention. When transport is via the public system, the reports will appear in the 911 data discussed in a previous section. An example of how these data can be used is found in Durham County of North Carolina at <http://www.durhamems.com/research.htm>.

Many medical emergencies are treated in emergency rooms but do not appear as 911 calls. A more complete measure of these emergencies could come from emergency room records. While there is no common database for these visits in most cities, there is a considerable interest in injury surveillance (Centers for Disease Control, 1988). The availability of E codes in the International Classification of Disease System makes it possible to establish data systems, and a growing number of communities are exploring such systems.

When many emergency rooms in a city collaborate on an injury registry system, it is possible to calculate injury rates for small areas. Among the important indicators are rates of intentional and accidental injury by age group (Rivara, Calonge and Thompson, 1989). Among youth in particular, injuries are a good indicator of health risk as well as social control in a community (Prothrow-Stith, 1991).

Immunization

The population's status on immunizations is considered an important measure of the adequacy of preventive health care. It is not only the protection afforded by the vaccine that is of interest but also the accompanying chance for a medical professional to examine the young child who is receiving immunizations according to the prescribed schedule.

There is no ordinance requiring states to collect data on immunization. Some states conduct a survey and estimate the number of children immunized, and some communities are experimenting with computerized immunization-tracking systems. Schools and day care centers are mandated to ask for proof of immunization before they admit students. Local area school registration as a source of immunization status at school entry was discussed in a previous section.

Some neighborhood indicators can be developed from available data. The immunization rate of children can be calculated to determine what percentage of children are not covered at time of entry into school or did not receive their vaccinations at the appropriate age. Incidence of communicable diseases for which there is a vaccine is another indicator of lack of immunization. More information about immunization can be accessed through the Centers for Disease Control and Prevention's immunization information website at <http://www.cdc.gov/nip/home.htm>.

Medicaid Claims

Medicaid provides medical assistance for certain individuals and families with low incomes and resources. Claims filed by medical providers for services delivered under Medicaid may be a valuable source of data on medical conditions and services for populations in small geographic areas.

Although Medicaid program administration varies from state to state, it is likely that Medicaid claims data can be obtained from the state agency that administers the program. Data may include provider description, classification of illness, procedure codes, service dates, and service charges. The recipient's address, needed for small area analysis, may need to be merged into the claim from an eligibility file. Most states enter only a limited number of variables into the computerized system. Owing to confidentiality issues, special requests justifying need may be necessary for release of data.

Several small area indicators such as annualized rates of types of medical care utilization (emergency, inpatient, ambulatory, etc.), annualized rates of Medicaid utilization by health status and age, and Medicaid utilization by type of service can be developed using such data. Statistical information about Medicaid at the national level can be viewed at the Health Care Financing Administration's website at <http://www.hcfa.gov/medicaid/mcaidpti.htm>. Additional national-level

Medicaid-related statistical information such as type of utilization can be found at <http://www.census.gov/ftp/pub/prod/1/gen/95statab/health.pdf>.

Claims data, which have been used extensively in health services research, are not likely to be available in the same format under managed care. Managed care providers may be required to submit encounter forms to the state for Medicaid enrollees. However, some states may accept aggregate reports of services provided for population groupings. This would preclude the type of address-based small area analysis being discussed here.

Hospital Discharge Files

Hospital discharge files contain information on hospitalizations generated at the time the patient is discharged. State hospital associations or government agencies may maintain such files. Although there are some common conventions for such data, they vary across locales.

Many state hospital associations maintain and publish data on patients aggregated at the zip code level. Data about age, payer, clinical service, sex, length of stay, diagnostic related group (DRG), hospitals, beds, and admissions may be available for small geographic areas such as a zip code. Patient-level data exist but are guided by confidentiality issues for release of information. The American Hospital Association publishes *Hospital Statistics*, which provides some hospital information with the address of the facility.

Some small geographic area indicators that can be developed are average charges by severity, number of inpatients and outpatients, incidence of most prevalent preventable condition per 1,000 population, average length of stay, number of beds in a small geographic area, and number of hospitals in a small geographic area. Researchers have compiled utilization rates across geographic areas using discharge data (Wennberg, Freeman and Culp, 1987). The National Association of Health Data Organizations (NAHDO) has addressed some of the data-related issues in its website at <http://www.nahdo.org/index.html>.

SOCIAL SERVICES

Social services are a range of public and private programs rendered to individuals and families to improve their economic, social, physical, and mental well-being. Service provisions and utilization are a useful indicator of the status of the population in small geographic areas.

Public Assistance Files

Various forms of cash and in-kind assistance are given to eligible persons who qualify under means-testing criteria. These programs operate under state and federal law but are delivered locally. Temporary Assistance for Needy Families has now replaced the well-known AFDC program. However, public assistance also includes food stamps, Medicaid, emergency assistance, and general assistance in many places.

Data on public assistance benefits are available through the state or local departments of human services. Computerized individual records including name and address, case and recipient numbers, program participation, eligibility status, and benefit amount are contained in monthly files. Records can be extracted for assistance units or for individual recipients. A few states maintain longitudinal records, but in many places these have to be created by merging monthly records to create a history for individuals.

Public assistance files are confidential and can be released only for valid purposes with proper protection agreements in place. Some departments have geo-coded their monthly files and can provide data with census tract codes rather than names and addresses, reducing the confidentiality problems. However, without recipient identifiers, longitudinal or matched files cannot be created.

Monthly files can be used to calculate participation in various public assistance programs for neighborhood residents. Longitudinal files can be used to calculate rates of long-term and short-term welfare participation. When public assistance records are merged with UI wage records (discussed above), rates of moving from welfare to work in small geographic areas can be calculated.

Subsidized Day Care

The child care programs operated under the Personal Responsibility and Work Opportunity Act maintain records on children receiving day care subsidies, their families, and day care providers. These data are confidential but may be available through state or local departments of human services for valid research purposes with proper protection. Records include name of parent and child, address, eligibility status, service hours per week, estimated cost per week, at-risk indicator, family size, income, education, and caretaker. The records are organized by month. Children receiving day care subsidy because of their participation in public assistance and those whose parents qualify because of their low-income working status are included. These records are confidential.

Subsidized day care participation of the families in neighborhoods can be calculated using these data. The amount of the subsidies can also be measured as well as the types of providers that are chosen. The research and publications department of the Department of Health and Human

Services has national-level statistics on various day care indicators at the website <http://aspe.os.dhhs.gov/GB/sec12.txt>.

Day Care Licenses

Every state has particular requirements for child care providers to obtain a state license. Directors and teachers who work in child care centers must have training in the care of children. Licensing rules also limit how many children can be cared for in any program and how many adults need to be present to care for children of certain ages.

Data on day care licenses are obtained through the state or local agency charged with the licensing responsibility. Local area child care resource and referral agencies often maintain computerized databases on child care providers as well. Confidentiality guidelines dictate the release of data. Requests justifying the need and guaranteeing protection for identifying information such as addresses can release such data.

Records contain the name and address of the licensee; number of slots for infants, toddlers, preschoolers, and school-age children; and additional child care information. Some indicators that can be developed for a small geographic area may include the numbers, type, and location of slots relative to employment locations and to the welfare and working poor population. Some information about day care can be obtained from <http://www.careguide.net/>.

Child Welfare Records

Child welfare services include preventive services such as social support, investigative services of reported child abuse or neglect, services for abused and neglected children, crisis intervention, and other such services. Child welfare records are maintained by county and state departments of child welfare or child protection. These records are confidential.

Although there are federal requirements for computerized child welfare information, the data systems vary in their completeness and accuracy. Records may include name, address, income, school, education, religion, ethnicity, marital status, and other demographic information. Dates and status of child abuse and neglect reports, entry and exit into custody, foster care, residential treatment, protective services, and special programs are also important pieces of information for analysis. In many states, the child abuse and neglect reports are in different data systems than those containing the ongoing protective services.

Neighborhood measures can be developed using geo-coded records. The number of children in custody or in foster care can be calculated for small geographic areas. Rates of reported child maltreatment can also be calculated. Child welfare records can be merged with other agency records to examine the interrelationships among services provided by various child-serving agencies and to

track outcomes (Goerge, Van Voorhis and Lee, 1994). Although such matching and merging are challenging, they have the potential to provide the kind of information useful to community organizations. The website at <http://aspe.os.dhhs.gov/hsp/cyp/chapin1.pdf> contains information on child welfare databases.

Mental Health, Alcohol, and Drug Services

Mental health, alcohol, and other drug abuse services are delivered by a wide range of public, nonprofit, and for-profit organizations. Although no single agency maintains data on all such programs, those operated under the authority of local boards often generate data useful for developing small area measures. However, the data are limited because many services do not fall under the public system.

Local boards often maintain confidential, computerized records that contain the client's name, address, and dates and types of service, including admissions to inpatient facilities or treatment centers. The availability and format of these records vary considerably from one local board to another.

Some community indicators can be developed from the geo-coded records. These include rates under treatment, demographic characteristics of patients, prior treatment history, criminal justice history, social support, addiction severity, duration of treatment episodes, key services received, program staffing, ownership, resource base, and costs.

The federal government has a Substance Abuse and Mental Health Services Administration (SAMHSA) that provides national-level information at <http://www.samhsa.gov/>.

SAFETY AND SECURITY

Safety and security within the neighborhood are of major concern in almost all communities. Information regarding the safety of a given geographic area can be found in several data sources.

Municipal Police

Police departments maintain records for each incident of reported crime occurring in their jurisdiction. These records contain a significant amount of information about the crime, the victim, and, when available, the suspect(s) and/or arrestee(s).

The crime reports contain specific information about the crime, such as location, type of crime, time, date, weather conditions, and information about the arresting officer(s) including name(s) and badge number(s). Types of crime include homicide, rape, aggravated assault, robbery, burglary, arson, auto theft, domestic violence, simple assault, menacing, and drug-related violations such as trafficking or possession. For a portion of the crime reports there is also a file of information about the victim, including race, sex, address, age, and date of birth. Police departments also maintain an arrest database. Included in each arrest report are address, race, sex, age, and date of birth of the arrestee. Information is also available about the suspect(s). It includes the same geographic and demographic information as the arrestee data but in addition there are physical characteristics describing a suspect. Information is also available regarding the type of weapon used during the crime. Crime reports can be linked to victim, arrest, and suspect reports using a report number. Most departments have data that are available annually and electronically. Since this information is confidential and sensitive, the willingness of a police department to release these data may vary by jurisdiction.

There are several small area indicators that can be developed with these data. The number and rates of crime by geographic area can be calculated. When making these counts, many researchers use only serious crimes, called Part I crimes under the terminology of the Uniform Crime Reports. Crimes can also be disaggregated by the race, sex, and gender of victims and assailants or by the victim-assailant relationship. For example, it can be determined whether the victim and assailant live in the same tract or whether they are of the same race. Weapons use by crime type can also be calculated.

Crime data have been used by a variety of groups and individuals, including researchers, planners, and administrators, to document and study crime levels in communities. Communities have used crime data to determine the need for violence-prevention programs or community policing programs. The relationship between crime and other social indicators has been studied to understand or determine the possible causes and effects of crime.

Most police departments have adopted the Uniform Crime Reporting (UCR) program codes for crime reporting and standardization of offense definitions. This standardization allows for comparison of crime statistics across police jurisdictions. The Federal Bureau of Investigation (FBI) administers the UCR program, which compiles and maintains these nationwide crime statistics. Local and state law enforcement agencies voluntarily participate in this program. The FBI annually releases a report entitled "Crime in the United States," which contains the data collected through the

UCR program. The report provides data by state, county, and for cities and towns with populations over 10,000. The report includes crime rates, number of crimes by type, and number of arrests by sex, race, and age. These annual reports can be found in libraries and can be purchased from the U.S. Government Printing Office. The FBI maintains a website at <http://www.fbi.gov>.

Emergency 911 Call Data

The 911 emergency system is generally operated by a county agency. All 911 calls for emergency service are processed by the appropriate police department. The calls for fire or EMS are routed to the appropriate agency.

The data available from the 911 calls include information about the caller and the emergency. Information about the caller includes name, address, and phone number. This information is not always complete, given a person's reluctance to identify himself. Available information about the emergency includes the exact location, date and time of call, description of emergency, whether an ambulance was required, and a priority and alarm level based on type of emergency. The time a call is received by police, time of arrival to the scene, whether a contact was made, and description of the result of the call are also available.

The data are organized by priority level, with 1 being the most serious and 4 being least. The data are categorized by type of call, such as whether the call indicates a crime against a person, an accident (e.g., hazardous waste), public safety (e.g., bomb threat), or a property crime or general assistance (e.g., hauling a prisoner, assisting with a traffic stop). The data are available annually. The ability to obtain these data and the format of the data may vary by jurisdiction.

The small area indicators that can be developed from these data include police response times and number of 911 calls by priority level. In addition, the calls can be categorized based on the description of the emergency. For example, the number of 911 calls indicating violent crime (e.g., homicide, robbery, domestic violence) can be calculated. In addition, the number of 911 calls indicating property crimes or public safety issues can be calculated.

The data from 911 calls can provide a supplement to other crime-related information. The 911 data provide information about incidents that may not result in a crime report, such as altercations within households or minor disturbances. The data regarding response times could be useful for police departments and communities as a whole. Slow response times may indicate a need for more staffing during particular time periods or in particular areas.

Juvenile Court

The Juvenile Court handles cases of delinquency, unruliness, abuse, neglect, and dependency for all individuals under the age of eighteen. A record is maintained for each juvenile who enters the court system.

The computerized Juvenile Court records often include information about the offender, including age, sex, race, the tract where the offender lives, and birthdate. There is information about the offense, including location where the offense took place, date of the offense, and type of offense. Offenses include violent crimes such as homicide and robbery, property crimes, drug violations, and less serious offenses such as disorderly conduct, curfew violations, and truancy. The data include information identifying the case such as a case number, case type, source of complaint (e.g., parent, school), judge, disposition, and disposition date. In addition, there is information regarding probation, such as probation officer, days on probation, and, where applicable, detention home location and release date.

Although the addresses of the offender and victim and location of offense are recorded, the ability to obtain this information will vary by court system. In addition, the ability to obtain demographic characteristics about the victim will also vary by court system. Detailed information about the victim and offender, such as address, race, sex, and age, is confidential. The Juvenile Court may release an annual report aggregating data to the municipality or county level, but neighborhood indicators are seldom published.

The community measures that can be developed from this include delinquency rates; number and type of crimes committed by juveniles; and the race, sex, and age of the offenders. If the victim information is obtained, the victim-offender relationship by sex, race, age, and geography can also be determined. The data have been used by a variety of individuals and groups, including researchers, policymakers, juvenile justice advocates, and administrators, to determine the level of juvenile crime and develop strategies and programs to reduce it. The Office of Juvenile Justice Prevention provides information at <http://www.ncjrs.org/ojjdp/html/pubs.html>.

Coroners' Reports

The coroner determines the circumstances, manner, and cause of all violent, sudden, unusual, and unattended deaths. The coroner prepares a detailed report outlining all of the findings surrounding a death. The coroner's report contains confidential information about the victim and, when applicable, the assailant. Information about the victim includes age, sex, race, and address. Additionally, there is information regarding any drug(s) or alcohol found in the victim's system at the time of death. Information about the death includes place, date, time, and day of death; type of death (e.g., homicide, suicide); mode of death (e.g., firearm, stabbing); and, where applicable, caliber of weapon. The information available about the assailant includes sex, race, age, address, and

previous offenses. The records are maintained by a local coroner's office. The data are available annually but the format and ability to obtain the data may vary by jurisdiction. Much of the report, for example, may not be computerized.

The community indicators that can be developed from these data include suicide and homicide rates, places of death, times and days when death occurs, and how death occurred. Information about the victim and assailant, such as race and sex, can be determined. In addition, the victim-assailant relationship by age, race, sex, and geography can be determined. Coroners' reports can be used in conjunction with police crime reports to enhance the information available regarding homicides and suicides. Coroners' reports contain more detailed information about a death than do the crime reports from a police department.

Child Maltreatment

Incidents of child abuse and neglect are reported to local child protection agencies. The agency investigates each claim and determines whether abuse and/or neglect has occurred or is occurring. A record is maintained for each reported incident of abuse and/or neglect.

The data available from reported cases of child abuse include location of alleged abuse and/or neglect; person reporting alleged abuse and/or neglect (e.g., teacher, doctor, victim); type of alleged abuse and/or neglect (e.g., sexual, physical, emotional); and whether abuse and/or neglect was substantiated, indicated, or unsubstantiated. Information about the victim is also available. It includes victim's address, age (which is calculated), and gender and victim's relationship to the perpetrator. Perpetrator information, including age and gender is available.

The confidentiality and sensitivity of this information is an important issue, particularly at the address level. The ability to obtain these data varies depending on the purpose for which they are sought. Generally, the data are available annually and electronically.

Indicators that can be developed from these data include child abuse rates, types of abuse and/or neglect being reported, and numbers of substantiated versus unsubstantiated incidents. The age and gender of victims can be determined. The perpetrators of abuse (e.g., parents, relatives) and those reporting the abuse (e.g., social workers, teachers) can also be determined. The data have been used by a variety of groups and individuals, including researchers, child welfare advocates, and policymakers, to document the problem of child maltreatment and determine the factors that may contribute to it.

The National Center on Child Abuse and Neglect (<http://caliber3.calib.com/nccanch>), part of the U.S. Department of Health and Human Services, collects and maintains national- and state-level information on abused and neglected children through its National Child Abuse and Neglect Data System (NCANDS), but it is not released for small areas. It is a voluntary program, but most of the

states provide data to the system. The information is compiled in an annual report and includes who is reporting the maltreatment; number of reports substantiated, unsubstantiated, or indicated; types of maltreatment; and perpetrator and victim characteristics such as age, race, and gender.

Liquor Licenses

A state's Department of Liquor Control is responsible, among other duties, for issuing permits to manufacture, sell, and distribute alcoholic beverages. It collects and maintains a record for each individual or company involved in one of these activities. This is public information. Records include name of permit holder and address of permit location, including street, city, zip code, county, taxing district, and permit class (e.g., carryout beer only, wine only). The address can be problematic in liquor permits. In some cases, the address may not represent the true location of the outlet. However, in these cases perhaps the name of the permit holder may provide some clue to its true location.

There are a few community measures that can be developed from these data. It can be determined how many alcohol outlets there are in a given geographic area and the number per capita. It can also be determined whether the outlets are carryout stores or bars. In addition, it can be determined whether these alcohol outlets carry beer, wine, hard liquor, or all three and how late the alcohol can be sold. The data have been used by researchers to examine and document the relationship between the density of alcohol outlets and the level of violent crime offenses in an area.

COMMUNITY RESOURCES AND PARTICIPATION

There is growing recognition that community participation and neighborhood resources are important to its residents. Further, community organizations are often interested in rebuilding community infrastructure and capacity. These phenomena are difficult to capture using administrative data, but a few sources are listed here.

Voter Records

Voter registration and participation are recorded by the local board of elections. These records are public but vary in their format and accessibility. Numbers of registered voters may be routinely reported by ward or other political jurisdiction that does not fit the boundaries of community organizations. Therefore, addresses are needed to calculate the rates of participation and registration that would be useful indicators of civic involvement. Voter records have been used to measure participation by neighborhood and ethnicity. For an example for Latinos, see <http://naid.spsr.ucla.edu/southwest/test1.html>.

Membership Records

Membership in neighborhood organizations is another indicator that may be useful. Organizations vary in the quality and format of their membership records. Those that collect dues and recruit and service their members are more likely to have up-to-date, computerized records. If individual addresses are available, these records can be geo-coded to obtain counts of members for small geographic areas. Increases in local membership may be used as indicators of rising social participation.

Public Transit

Access to public transportation and the destinations accessible by public transport are an important indicator for small areas. It is particularly important that nondrivers such as the elderly, working and welfare poor, and people with disabilities can reach important locations within the metropolitan area and can reach transit stops within their neighborhood.

Bus and train schedules and routes are published by local transit authorities and can be converted to travel times to key destinations from points of origin within neighborhoods. Some transit authorities may have computerized route and schedule information that can facilitate this task. Transit stops can be geo-coded, mapped, and used in calculations of residents' distances to stations.

Small geographic area indicators can be developed from geo-coded transit data. The average time of commute to areas of employment growth is an example that is important to economic development (Coulton, Verma and Guo, 1966). With respect to service reform, average commute to service providers might be another useful indicator. Some national information is available on the American Public Transit Association website <http://www.apta.com/>.

Auto Registration and Licenses

An additional indicator of a neighborhood's access to the region and its general level of resources is auto registration and licensed drivers. Records of auto registration and licenses are public and can be obtained for a fee from the state's Bureau of Motor Vehicles. Some states use a dynamic database that is updated when changes are made. The data available are a snapshot at a given point in time. Data are available with addresses and zip codes for the various types of vehicles registered and the classification of the vehicles (commercial, passenger, government, etc.). The address field can be used to assign a census tract using existing geo-coding software. Examples of community measures that can be developed using such data are the number of registrations done in a year, the number of automobiles owned per capita and the average number of vehicles owned by a household.

Community Development Block Grants

The Community Development Block Grant (CDBG) program is funded by the Department of Housing and Urban Development (HUD) and administered by the community receiving the funding. It provides important resources for neighborhoods. This entitlement program provides annual grants to central cities of Metropolitan Statistical Areas (MSAs), other metropolitan cities with populations of at least 50,000, and qualified urban counties with populations of at least 200,000. The purpose of these grants is to assist communities in carrying out a wide range of community development activities directed toward neighborhood revitalization, economic development, and the provision of improved community facilities and services. Specifically targeted are those areas where there is a concentration of persons with low and moderate incomes. Each city allocates the funding to projects it deems appropriate and consistent with HUD regulations.

These data are prepared and maintained by the individual community receiving the funding, and more than likely by the community's department of community development or planning. The data available from CDBG include name and address of the funding recipient, census tract and political boundary of recipient, description of activity funded, activity codes, amount of funding, month and year activity was funded, amount of the funding expended in a given period, geographic area served by the activity (census tract(s), political boundary, or city as a whole), and number of persons and/or households served by the activity. Activities include housing rehabilitation, new construction, public facilities, and improvements such as neighborhood centers, parks, street improvements, and health facilities. In addition, the racial and income characteristics of the persons and/or households served by particular activities are also available. There are codes (called National Objective Codes) that indicate whether the area being served is considered low/moderate income, slum/blighted, or in urgent need of assistance.

Even though this information is public record, the ease of obtaining and using the information may vary by community. Each city must submit an annual report to HUD outlining the previous year's CDBG activities. Whether the data are available electronically may vary by community.

There are some ambiguities with these data, particularly when the user is focusing on neighborhood analysis. It is difficult to accurately determine the financial benefit a given neighborhood is receiving from some types of CDBG activities. Some projects, for example, target multiple neighborhoods. Or a project located in a neighborhood may offer city-wide programs. Such expenditures must either be apportioned across many neighborhoods or left out of neighborhood-level analyses.

There are some community indicators that can be developed from these data. It is possible to estimate the CDBG funding allocated by geographic area. An estimate of the per capita funding by geographic area can be calculated. It is also possible to determine what types of activities are being funded by geographic area. The racial composition of those being assisted can be calculated.

In addition, the income categories (i.e., low/moderate or low) of those being assisted can be determined. The data have been used to study the impact CDBG funding has on residential rehabilitation and other community development activities in specific communities. The data have also been used to determine the investment being made in the various CDBG activities.

Community Directories

Organizations are among the assets of neighborhoods that are important to planning. Unfortunately, there is no single database that provides information on all of the community-based organizations. Nevertheless, there are some listings that may be useful for this purpose. The computerized yellow pages is a source for organizations that are listed there. The yellow page addresses can be geo-coded and mapped by neighborhood. Information about the computerized yellow pages can be obtained from <http://www.nctweb.com/CDS/selphone.html>.

Many organizations belong to associations whose directories may be useful in constructing neighborhood measures. For example, a listing of libraries and their addresses can be obtained from the American Library Association. Churches may be listed by a local interchurch council. Neighborhood development corporations and neighborhood centers may be listed by the local economic development agency. Parks and playgrounds are another part of community resources that can be mapped using information from the local parks administrative department.

While available listings can provide the geographic location of community assets, the listings seldom provide data that can capture the magnitude of the organizations' operations or contributions. Such information may need to be obtained from original data collection directly from the organizations themselves.

HOUSING

Many community organizations are interested in improving the housing in their neighborhoods. There is a considerable amount of data available that can be used to develop indicators of housing stock, conditions, and markets in small areas. The data come from a variety of sources. Information available includes the characteristics of housing, the condition and quality of housing, the construction and demolition of housing, and the financing of housing.

Home Mortgage Disclosure Act (HMDA)

The Home Mortgage Disclosure Act (HMDA), enacted in 1975, is implemented by the Federal Reserve Board. This act requires covered institutions to compile and disclose data about loan applications they receive and home purchase and home improvement loans they originate or purchase during each calendar year. Institutions that are required to file HMDA data include

commercial banks, savings and loans, credit unions, and mortgage companies that meet specific reporting criteria.

The data are maintained in the institution's Loan Application Register (LAR). Each LAR record contains loan application information such as type, purpose, amount, and action taken. Each record also contains some applicant and co-applicant characteristics such as race, gender, and gross annual income. Information about the property location, such as the census tract, county, MSA, and state, is also in each LAR record. In addition, each record includes a few census-tract-level variables such as population, number and percentage of minority population, median income, and number of owner-occupied units.

In addition to the LAR record, there is a Transmittal Sheet (TS) record. The TS record contains information about each financial institution, such as name, address, parent company name and address, and tax identification number. The LAR and TS data can be linked by using a respondent ID and agency code found on both of these files.

These data are provided on reel, cartridge, and CD-ROM by the Federal Reserve Bank's Federal Financial Institutions Examination Council (FFIEC). The data are available annually and the current year's data are released in the fall of the following year. The order form can be requested by calling the FFIEC or downloaded from the council's website at <http://www.ffiec.gov/hmda>.

There are several community measures that can be developed from the HMDA data. The total number of loans applied for and whether they were approved, denied, or withdrawn can be determined. The reasons for denial can be determined. The purpose for the loan/application, such as whether it was to purchase, improve, or refinance a home, is available. The type of loan/application—such as whether it was conventional, Federal Housing Administration (FHA), Veterans' Administration (VA), or Farmers' Home Administration (FmHA)—is available. Demographic characteristics about the loan applicants and co-applicants is available. Approval and denial rates based on race, gender, and income can be computed. The approval and denial rates of the financial institutions can also be computed for small areas. The financial institutions doing most of the lending in a particular area or to a particular group can be determined, as well as the lending patterns of specific financial institutions.

The data have been used by fair-housing groups to look at lending patterns in communities and determine whether any discriminatory practices are involved. The data are also useful for public officials to determine whether financial institutions are meeting the housing credit needs of their communities. Additionally, the economic stability of neighborhoods can be assessed by computing the trends on commercial lending and conventional mortgage activity.

Property Data from Local Tax Assessor or Auditor

A wide variety of information about every parcel of property is collected and maintained by the local auditor or assessor office. This information is collected for the purpose of levying taxes.

There are three types of data about a property. The tax billing record includes parcel number, parcel size, owner's name and address, land and building assessed values, property class, land use codes, gross taxes, special assessments, and delinquency status. The characteristics data include parcel number, number of rooms, year built, lot size, land use code, and roof type. The deed transfer data include information about property sales and transfers, names of buyer(s) and seller(s), address of property, sales amount, date of sale, and deed type.

While this information is public record, the ease in obtaining it in a usable format for analysis may vary from one region to another. Some local offices may be able to provide the data electronically while others may not have the resources to do so. All three types of data can be linked by parcel number. The tax billing and characteristics data are available annually. The deed transfer data are available monthly.

Some of the same property information is contained in a commercial software product from Transamerica Information Management called MetroScan. It is intended for use by realtors. The MetroScan database includes the census tract and block code for the parcel as well as some school district information. In addition, the user has the ability to print county and street maps. The MetroScan database can be searched by a number of variables in the database, including tract, street, and property type. Although the data can be exported in several formats, the user can export only 5,000 records at a time. This would be very cumbersome for areas with a large number of properties but would be convenient for a modest-size neighborhood. MetroScan is available for approximately 130 counties nationwide; information can be obtained at <http://www.transamerica.com>.

There are several small area measures that can be developed from real property data. The market and assessed values of a home can be computed. Median and average sales prices can be calculated. The number of tax delinquent properties can be determined. Volume of property sales and transfers can be determined. Number of sales by deed type (e.g., sheriff, trustee, warranty) can be determined. All of these indicators can be computed by property class (i.e., residential, commercial, industrial), land use (e.g., single-family, commercial warehouse, manufacturing plant), and geographic area.

The data have been used to study small area trends in housing values and the market for real estate. Also, studies have examined the pattern of population movement by linking deed transfers for sale and purchase, thus tracking where individuals are buying and selling property. Many of these indicators, in conjunction with other indicators, have been used to assess neighborhood stability and condition.

Building and Demolition Permits

Building and demolition permits are collected and maintained within a municipality's building or housing department. The permits are necessary to ensure that zoning requirements, fire and structural standards, and other building standards are met.

The data available on permits include parcel number, name of owner, address of property, written description of work to be done, codes identifying work to be done, permit use class (e.g., commercial, residential), estimated cost of work to be done, permit issue date, and expiration date. In addition, there is information about the contractors performing the work, such as name and registration number. The building permit also contains the inspection date, building inspector, and permit fees. The information is public record and can be obtained in the appropriate city department. However, the ability to obtain the data in an electronic format may vary by city.

The small area indicators that can be developed from this data include the number of permits by type (e.g., new construction, external rehabilitation) and the number of permits for commercial versus residential properties by geographic area. Additionally, it is possible to calculate the value associated with the building permits by type and geographic area. The data have been used to determine the investment being made in residential rehabilitation and new housing construction by geographic area. Also, neighborhood groups have been interested in tracking housing demolition. In addition, the data have been used, in conjunction with other housing indicators, to assess neighborhood stability and condition.

Housing Code Enforcement

Housing code enforcement information is available from the city's building inspection or housing division. The housing code sets the rules for basic maintenance and upkeep for decent housing. The main purpose for housing codes is to ensure the health and safety of the occupants of a building.

The information available from a housing code violation report includes parcel number, address and owner of property with violation, code being violated, inspector name, date of violation cited, time period allotted to comply, compliance date, and whether legal action was taken. In addition to code violations are nuisances, which are problems that make a neighborhood unattractive or unsafe.

The information is public record and can be obtained from the appropriate city department. However, the ability to obtain the data in an electronic format may vary by city. The data are frequently updated.

The small area indicators that can be developed from this data include number and rate of properties with violations, types of violations (e.g., faulty wiring, paint needed), major versus minor violations, and number and rate of nuisances (e.g., abandoned buildings and cars, garbage improperly stored) by geographic area. The compliance rates of code violations or nuisances can be calculated. In addition, the number of violations that result in legal action can also be calculated.

The data have been used, in conjunction with other housing indicators, to assess the stability and condition of neighborhoods. In addition, the data have been used to document a community's success in enforcing compliance with code violations.

Public and Subsidized Housing

The Department of Housing and Urban Development (HUD) collects and maintains a variety of information about public and subsidized housing units across the nation. There is information available about the tenants of public and subsidized housing and about the housing projects. Generally, the information is collected at the local level and reported to HUD.

The data from HUD are available at the census tract level and individual housing project level. The information includes the name and address of the housing project, total number of subsidized housing units by type (e.g., public housing, Section 8), percentage of units occupied, average rent, percentage of tenants who moved during the past year, average stay in unit, average number of months on waiting list, average number of persons per unit, average income, percentage of persons in different income categories, where majority of tenants' income comes from (percentage from welfare and percentage from wages), percentage with assets over \$5,000, average age of head of household, percentage with disability, racial breakdown, average bedroom sizes, percentage of persons by family type, percentage overhoused (more bedrooms than people), percentage with utility allowance, average stay, and average months on waiting list. Geographic information includes zip code, average latitude and longitude, census tract surrounding the housing project, and county. Additional tract-level information includes percentage poor, percentage minority, and percentage of households that are owner-occupied. This information is available through HUD's HUDUSER database by project site and/or the census tract in which the site is located. The website is <http://www.huduser.org/data.html>. The data can be downloaded for inclusion in databases, spreadsheets, and statistical packages.

Each local housing authority collects and maintains its own records regarding its tenants and/or management of the units. The locally maintained databases may provide more detail and flexibility than the national data set, which provides only aggregate, rather than individual, information. Also, because actual addresses are available in local housing authority data, the geocoding can be more complete than that currently available through HUD. However, the ability to obtain information may vary by housing authority.

Public housing data can be used to determine the economic status and mobility of public housing residents within the neighborhood. Combined with a total housing unit count from the census, it can be used to calculate the percentage of housing in the neighborhood that is publicly operated or subsidized. The data have been used by a variety of researchers, evaluators, and housing administrators to profile the public and subsidized housing population and develop programs to assist tenants in moving from subsidized housing to home ownership.

THE ENVIRONMENT

While environmental data have not been widely used by community groups, there is a large amount of information that is now available for examining the environmental quality of the neighborhoods in which people live. The largest single source of such data is the U.S. Environmental Protection Agency (EPA), which maintains several different databases on air quality, compliance with environmental regulations, release of toxic materials, pollution discharge, and Superfund sites. Because of EPA's role as a regulation-enforcing organization, most of the data relate to monitoring sources of pollution rather than to overall environmental quality.

EPA's databases are maintained for different purposes by separate program offices within the agency. Up until now, these data systems have not necessarily been compatible with each other. At this time, however, EPA is making an effort to integrate all of its data geographically by collecting accurate locational information on different entities of environmental concern (such as monitoring stations and factories). The intent is to promote the use of EPA's data resources for a wide array of cross-media analysis, such as community-based ecosystem management and environmental justice. EPA's goal is to obtain latitude and longitude coordinate information (with plus or minus 25 meters accuracy) for all regulated facilities and sites, operable units, and environmental monitoring and observation locations by the end of the year 2000.

The first result of these efforts is EPA's Envirofacts Warehouse, which integrates data from eight separate program databases through an easy-to-use interface. Through the Envirofacts website (http://www.epa.gov/enviro/html/ef_home.html), users can query the different databases to obtain information on sources of air pollution, compliance with water-discharge permits, Superfund sites, releases of toxic chemicals, hazardous waste sites and transporters, and over 675,000 EPA-regulated or monitored facilities.

Envirofacts permits users to limit their searches to locations in a particular state, city, county, or zip code. The locational information in the database includes the street address, city, state, county, and zip code, as well as the latitude and longitude for each site. With the latitude and longitude, a user can map the locations of different sites using mapping software. While the EPA data do not include census geographic identifiers (e.g., tract and block group), the user can add these by geo-coding the address or using the latitude and longitude to locate the specific site.

The Envirofacts website also includes a “Maps On Demand” function that allows the user to produce customized maps of data in the system.

The different types of data available through Envirofacts are described below. In some cases, it is possible to obtain additional or more detailed information by accessing the program database directly, as opposed to using the Envirofacts website. We note these situations below and describe how to go about obtaining direct access to these other sources.

Air Pollution

The Aerometric Information Retrieval System (AIRS) is maintained by EPA’s Office of Air Quality Planning and Standards (OAQPS) and contains data about airborne pollution in the United States and in various World Health Organization member countries. The AIRS database actually consists of three subsystems—the AIRS Facility Subsystem (AFS), the Air Quality Subsystem (AQS), and the Geographic, Common, and Maintenance Subsystem (GCS). Only the first of these (the AFS) is included in the Envirofacts Warehouse.

The AFS contains both emissions and compliance data on about 150,000 air pollution point sources regulated by the EPA and/or state and local air regulatory agencies. AFS contains data on industrial plants and their components: stacks, the points at which emissions are introduced into the atmosphere; points, the emission points or processes within a plant that produce the pollutant emissions; and segments, which are components of the processes that produce emissions. Compliance data are maintained at the plant and point levels, tracking classification status, inspections, and compliance actions. AFS also includes data for management of operating permit applications and renewals.

The information in AFS is used by the states in preparation of State Implementation Plans (SIPs) to track the compliance status of point sources with various regulatory programs and to report air emissions estimates for pollutants regulated under the Clean Air Act. General source identification information is maintained, such as name, address, industrial classification, operating status, and description, as well as descriptive and parametric data on stacks, emission points, and processes within the facility.

Data from AFS can be used to locate the sources in a neighborhood of different types of air pollutants regulated under the Clean Air Act, including carbon monoxide, nitrogen dioxide, sulphur dioxide, lead, and particulates. Information is also available on the amount of emissions each of these sources produces annually, and whether or not the facility is in compliance with government clean air programs. This information can be used to identify the important air polluters in a neighborhood and to see how many are complying with emission standards.

While the AFS maintains data on sources of air pollution, it does not have any information on levels of pollutants found in the air. That information is located in the AQS, which contains measurements of ambient concentrations of air pollutants and associated meteorological data from thousands of monitoring stations operated by EPA, national, state, and local agencies. The information in AQS can be obtained only by accessing the AIRS system directly. EPA maintains a website for the AIRS system at <http://www.epa.gov/airs/airs.html> that explains the different ways that one can access AIRS. Since there are only about 10,000 monitoring stations in the entire United States, these data may not be suitable for neighborhood-level analysis. For instance, the city of Washington, D.C., has only three monitoring stations.

AIRS data may be accessed directly by anyone with an IBM-3270 terminal and an account on the EPA's mainframe. (Account requests are sent to OAQPS.) Alternatively, persons with access to the Internet can use an IBM-3270 terminal emulation program, as opposed to an actual IBM terminal, to reach the EPA mainframe. This method is explained in detail at <http://www.epa.gov/airs/internet.html>. EPA has recently initiated a reengineering project for AQS. In addition to other modifications to the system, the plans are to make AQS accessible to users through World Wide Web browsers rather than through terminals.

Another alternative is a Windows-based software package made available by EPA called AIRS Executive, which contains a select subset of information from the complete AIRS database. With AIRS Executive, users can download monthly updates of AIRS data and use their own Permit Compliance System (PCS) to produce reports on air quality levels from monitoring stations and emissions and compliance from regulated facilities. The AIRS Executive software and data files can be downloaded from the Internet at <http://www.epa.gov/airs/aexec.html>.

Finally, it is possible to request AIRS data from EPA through the Freedom of Information Act (FOIA). There may be charges for certain requests. A sample letter for a FOIA request is provided at <http://www.epa.gov/airs/foia.html>.

Water-Discharge

The Permit Compliance System (PCS) is maintained by EPA's Office of Enforcement and Compliance Assurance (OECA) and contains data on more than 75,000 water-discharge permits regulated under the National Pollution Discharge Elimination System (NPDES) of the Clean Water Act. The NPDES permit program regulates direct discharges from municipal and industrial wastewater treatment facilities that discharge into the navigable waters of the United States. The database includes information on permit issuance, permit limits, monitoring data, and other data on facilities.

The location of different wastewater discharge point sources can be identified and located with the PCS. Users can also identify those sources that are in violation of discharge regulations and are undergoing enforcement actions.

The Envirofacts Warehouse is the easiest way to obtain data from the PCS. Alternatively, extracts from the PCS are available on 3.5-inch diskettes. For more information, contact the National Technical Information Service in Springfield, Virginia, at (703) 487-4650.

Superfund Sites

The Comprehensive Environmental Response, Compensation, and Liability Information System (CERCLIS) is the official repository for site- and non-site-specific Superfund data in support of the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA). CERCLIS contains information on hazardous waste site assessment and remediation from 1983 to the present, including data on active sites from point of discovery to listing on the National Priorities List through completion of remedial and response actions. CERCLIS information is used to report official Superfund accomplishments to Congress and the public, assist EPA regional and headquarters managers in evaluating the status and progress of site cleanup actions, track Superfund Comprehensive Accomplishments Plans (SCAPs), and communicate planned activities and budgets.

CERCLIS data can be obtained through the Envirofacts website. Users can get the address, latitude, and longitude of Superfund sites; the type of incident that led to the site being classified as a Superfund site; and the current enforcement status and activity. These data can be used to identify Superfund sites that are in or near a neighborhood and identify those sites that are undergoing clean-up.

As alternatives to obtaining data through Envirofacts, CERCLIS summary data sets and standard FOIA reports are available free on diskette by calling the Superfund automated phone system at (800) 775-5037. Standard CERCLIS reports are also available at the Superfund website <http://www.epa.gov/superfund>.

Toxic Chemicals

EPA's Office of Prevention, Pesticides, and Toxic Substances maintains the Toxic Release Inventory System (TRIS). TRIS has data about the release and transfer of more than 300 toxic chemicals and compounds by medium of release (air, water, underground injection, land disposal, and off-site), reported by over 33,000 submitters. TRIS information is available for 1987 through 1995.

TRIS stores release-transfer data hierarchically by facility, by year and chemical, and by medium of release. TRIS also stores treatment and source-reduction data. At the facility level, TRIS stores facility name, address, latitude/longitude, and parent company. At the chemical level, TRIS stores Standard Industrial Classification (SIC), EPA identification numbers (EPA ID), and pollution prevention data (e.g., recycling, energy recovery, treatment, and disposal). At the medium level, TRIS stores names and addresses of off-site transfer recipient facilities.

Users can identify locations of toxic chemical releases in or near particular neighborhoods, going back to 1987. The amount of each release can be obtained, as well as whether the release occurred in the air, water, underground, or on land. TRIS also identifies the owner of the facility responsible for the release and the type of treatment used to deal with the release.

In addition to accessing the data through Envirofacts, users can access TRIS data by contacting the National Library of Medicine, Bethesda, Maryland, at (301) 496-6531.

Hazardous Waste

Under the Resource Conservation and Recovery Act (RCRA), generators, transporters, treaters, storers, and disposers of hazardous waste (as defined by the federally recognized hazardous waste codes) are required to provide information concerning their activities to state environmental agencies, who in turn provide the information to regional and national EPA offices.

The Resource Conservation and Recovery Information System (RCRIS) is used by the EPA's Office of Solid Waste and Emergency Response (OSWER) to support its implementation of the RCRA. RCRIS is primarily used to track handler permit or closure status, compliance with federal and state regulations, and cleanup activities. Other uses of the data include program management, regulation development, waste-handler inventorying, corrective action tracking, regulation enforcement, facility management planning, and environmental program progress assessment.

RCRIS allows the user to identify the locations of toxic waste facilities and indicates whether any corrective actions have been taken by EPA or other environmental agencies at these locations. Information about opening and closing of toxic waste sites is also provided by RCRIS.

RCRIS data can be accessed through Envirofacts. In addition, copies of RCRIS reports are available in hard copy from EPA or can be obtained through the EPA gopher site at gopher://gopher.epa.gov.

ISSUES IN USING ADMINISTRATIVE DATA

Using administrative agency data to make measures on small areas presents a series of challenges. How successfully these problems can be overcome depends upon the type and source of the data and specific local circumstances. Some general issues will be reviewed that may or may not be a barrier in specific situations.

Geographic Boundaries

The use of available data sources may present some limitations on geographic definitions. For confidentiality purposes, some data sources contain only census codes or administrative districts rather than individual addresses. It is desirable, though, to always obtain data at the smallest geographic unit available, such as the address or parcel level, since this allows maximum flexibility for the community organization to set and change boundaries and still make relevant calculations.

If these measures built upon available data are to be a measure of change, it is also important that the geographic boundaries used for data aggregation are consistent over time. Census boundaries change each decade, but administrative agencies that include census tracts in their data may be relying on the older definitions or use boundary files that have not been updated. Users of the data need to evaluate the compatibility of geography between data sources and across time.

Confidentiality

Many administrative data sources contain individual information that is protected either by law or custom. Unless the database already contains census block or tract codes, though, confidential information about individuals' street addresses must be released for small area analysis. Even census-tract-level aggregated information may breach confidentiality if there are only a handful of cases that fall into a particular category and they could be identified.

Administrative agencies can enter into confidentiality protection agreements with researchers who have a valid purpose for using the data to develop community measures. Users need to follow standard methods for guarding the data that have identifiers, for ensuring that only necessary and secure personnel have access to the data, and for guaranteeing that the confidential data will not be released. If researchers are located in institutions with federally approved institutional review boards, they should have their confidentiality protection methods reviewed by these bodies. Some

agencies have well established guidelines for confidential data while others have little experience in this area. However, with the exception of a few agencies that are strictly prohibited from the release of confidential data by law, most agencies can release the data if the users are made agents of the agency and agree to abide by agency rules.

The community measures based on confidential data must be calculated on areas that are large enough so as not to reveal any individual identities. There can be a problem when events are rare or when indicators are based on categories that are small. For example, employment by industry in a small area could be made up of just one or two large firms that could be identified. Thus, such figures could not be released.

The confidentiality issue requires a judgment of the risk-benefit ratio. The research institution needs to be judged for its reputation and expertise in the protection of human subjects in order to determine the risk that an inadvertent breach of confidentiality may occur. The agency also needs to weigh the benefit to itself and the community of having small area data available. Many agencies do not have the internal resources to look carefully at their data by small area. Thus, the benefits of releasing the data (under strong and binding confidentiality protection) for geographically based analysis are often significant if the agency is assured that information will be provided to them in return.

Data Accuracy

Small area information produced from administrative records suffers from four types of accuracy problems. The first has to do with inaccuracies in the records themselves. Data elements that are essential for the agency's work or subject to quality control are likely to be most accurate. Thus, for example, public assistance payroll records that stem from the issuance of a check are more accurate than intake information, such as educational attainment, which has no bearing on eligibility. An important issue for small area analysis is the accuracy of the addresses for the event that is being analyzed. A common problem is agency databases that overwrite original addresses that are pertinent to an event with address changes that occur after the fact. Users of administrative data for research purposes need to check with the agency about each data element to make a judgment about accuracy.

A second problem in accuracy is reporting bias. This arises in situations where an event must be reported in order for an administrative record to be generated. For example, crimes are known to be underreported to the police (O'Brien, 1985), and law enforcement jurisdictions differ in their response to crime reports (Sherman, 1989). These two factors can affect whether a crime record is generated and how the crime is classified. Child abuse and neglect reports are vulnerable to similar problems resulting from the biases in reporting and agency response (O'Toole, Turbett and Nalpeka, 1983).

The third issue has to do with the small number of rare events in small areas. Infant deaths, for example, will occur in very small numbers in a small area for a given year. A change in even one death can raise the infant death rate markedly without reflecting a true change in health status of the population or quality of health services. For rare events, small area indicators need to use multiyear averages or group neighborhoods together to achieve a large enough number so that there is confidence in the estimate. The literature on sample size and accuracy should be consulted (e.g., Lemeshow, Hosmer, Klar and Lwanga, 1990).

A final source of inaccuracy in developing measures from administrative data has to do with population estimates for small areas. Many indicators are calculated as rates in order to make them comparable across small areas that differ in size. For example, crimes are converted into a crime rate by dividing the number of crime reports in a year by the population. The decennial census is typically used as the best count of the population, but these numbers quickly become out of date. It is preferable to use population estimates for the years between censuses, but there are well-known problems with accurately estimating the population in small areas (Smith and Cody, 1994). Thus, many measures are made even more inaccurate by errors in both the numerator and denominator.

Data Extraction and Management

Administrative records come in many different formats. Most easily usable by community groups are those that have already been aggregated to the units of geography that are of interest. Under these circumstances, the data file may contain one record for each census tract or neighborhood, and the required measures are easily extracted. For example, many local health departments routinely produce counts of births, deaths, and infant deaths and other useful statistics by census tract.

Most data, however, have not been converted to this format by administrative agencies. Typically, the data files contain individual-level records. The information for one individual may appear in multiple records and the records may be spread across multiple files. Some data files are extremely large and contain cases and records that are not of interest for calculating the small area measures. Considerable work often needs to go into understanding the file formats, extracting the relevant records, geo-coding the addresses, and aggregating the data to the required units of geography.

A complicating factor in using administrative data is ensuring that the correct records have been extracted for the measure that is desired. Although not always explicit, the measure requires that decisions be made about (1) a "window" of time; (2) whether persons or events are the unit; (3) whether all cases or just new cases should be counted; and (4) how to handle duplicates. For example, a child maltreatment report is an event that involves one or more children. In a given year, the same child may be reported multiple times. Further, a single event may yield several reports. Child maltreatment cases may be carried as open records in the agency database over several

months or years. Such data make possible several different measures for a small geographic area. Among the possibilities are counts of (1) the total number of maltreatment reports in a year; (2) the number of individual children who were reported as maltreated at least once in a year; (3) the total number of maltreatment cases served by the agency at a point in time during the year; and (4) the total number of maltreated children ever served during the year by the agency. In these respects, users need to be clear about exactly how their calculations are made and what the resulting measures mean.

Matched and Longitudinal Files

Administrative data are often organized by month, quarter, or year. Most are event driven, generating a record, for example, when a person is eligible for a program, a payment is made, a deed transfers, or a child is born. However, communities may require some measures that reflect the fact that these events happen over time to individuals, houses, firms, or some other entity. To develop this type of measure, longitudinal records must be created from the multiple events. This requires matching across separate records using a constant identifier such as a parcel number or a case number.

Although longitudinal measures require considerably more processing effort, they are frequently more reflective of important outcomes. For example, a community organization may be less interested in reducing public assistance use than in eliminating long-term dependency. The number of long-term welfare recipients in a small area can be calculated only if longitudinal files are created by matching the monthly eligibility files for each individual over a set number of years.

A similar challenge occurs when measures require that data be matched across multiple agencies or multiple data sets from a single agency. For example, a community organization may be interested in ensuring that preschool children of mothers moving from welfare to work are still able to take advantage of Head Start programs in the neighborhood. A match would need to be made among public assistance, employment, and Head Start records to monitor progress on this outcome.

Matching across multiple agencies may require the use of probabilistic matching procedures because there is no universal or reliable individual identifier (Jaro, 1995; Newcombe, 1988). For example, social security numbers are erroneous or not available in many data files but names, birthdates, addresses, and other identifiers can be used to improve the accuracy of matching. Matching across databases has the potential to create some of the most sensitive and refined measures that would be most useful in capturing the synergistic effect of comprehensive community initiatives (Goerge, Van Voorhis and Lee, 1994).

ENDNOTES

1. This catalog was largely written by Claudia J. Coulton, assisted by Lisa Nelson. Their material is adapted from a book chapter originally prepared for the Evaluation Subcommittee of the Aspen Institute Roundtable for Comprehensive Community Initiatives. Peter Tatian of the Urban Institute contributed sections on environmental data sources. The catalog was published in this form in G.T. Kingsley, C. Coulton, M. Barndt, D. Sawicki, and P. Tatian, *Mapping Your Community: Using Geographic Information to Strengthen Community Initiatives* (Washington DC: U.S. Department of Housing and Urban Development, October 1997).

REFERENCES

- Annie E. Casey Foundation. 1997. *City Kids Count*. Baltimore, MD: Annie E. Casey Foundation.
- Carlson, V. (1995). Identifying neighborhood businesses: A comparison of business listings. *Economic Development Quarterly*, 9, 50-59.
- Centers for Disease Control (1988). Public health surveillance of 1990 injury control objectives for the nation. *Morbidity and Mortality Weekly Report*, 37, 27-68.
- Coulton, Claudia J., Verma, N., and Guo, S. 1996. *Time-Limited Welfare and the Employment Prospects of AFDC Recipients in Cuyahoga County*. Cleveland: Center for Urban Poverty and Social Change, Case Western Reserve University. December.
- Geronimus, A. T., Bound, J., Waidmann, T. A., Hillemeier, M. M., and Burns, P. (1996). Excess mortality among blacks and whites in the United States. *New England Journal of Medicine*, 14, 491-498.
- Goerge, R., Van Voorhis, J., and Lee, B.J. (1994). Illinois longitudinal and relational child and family research database. *Social Science Computer Review*. 12, 351-365.
- Gould, J. B., Mahajan, N., and Lucero, M. (1989). Improving perinatal outcome through data management: The design of small area analyses system. *Journal of Perinatal Medicine*, 16, 305-314.
- Jaro, M. A. (1995). Probabilistic linkage of large public health data files. *Statistics in Medicine*, 14, 491-498.
- Kessner, D. M., Singer, J., Kalk, C. E., and Schlesinger, S. (1973). *Infant death; An analysis by maternal risk and health care*. Washington, DC: Institute of Medicine and National Academy of Sciences.
- Leete, L., and Bania, N. (1995). *Assessment of the geographic distribution and skill requirements of jobs in the Cleveland-Akron metropolitan area*. Cleveland, OH: Case Western Reserve University.
- Lemeshaw, S., Hosmer, D. W., Klar, J., and Lwanga, S. K. (1990). *Adequacy of sample size in health studies*. John Wiley & Sons Ltd.

- McCord, C., and Freeman, H. P. (1990, January 18). Excess mortality in Harlem. *The New England Journal of Medicine*, 322 (3), 173-177.
- Newcombe, H. B. (1988). *Handbook of record linkage: Methods for health and statistical studies, administration and business*. Oxford: Oxford University Press.
- O'Brien, R. M. (1985). *Crime and victimization data*. Beverly Hills, CA: Sage Publications.
- O'Toole, R., Turbett, P., and Nalpeka, C. (1983). Theories, professional knowledge, and diagnosis of child abuse. In D. Finkelhor, R. Gelles, G. Hotaling and M. Strauss (eds.), *The Dark side of families: Current Family Violence Research*, (pp. 349-361). Newbury Park, CA: Sage Publications.
- Prothrow-Stith, D. 1991. *Deadly Consequences*. New York, NY: Harper-Collins.
- Rivara, F. P., Calonge, N., and Thompson, R. S. (1989). Population-based study of unintentional injury incidence and impact during childhood. *American Journal of Public Health*, 79, 990-994.
- Rossi, P.H. (1972). Community social indicators. In A. Campbell and P. E. Converse (eds.) *The human meaning of social change*. New York: Russell Sage.
- Schorr, L. B. (1994). *The case for shifting to results-based accountability with a start-up list of outcome measures with annotations*. Cambridge, MA: Harvard Project on Effective Services.
- Sherman, L. (1989). Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology*, 27, 27-56.
- Smith, S. K., and Cody, S. (1994). Evaluating the Housing Unit Method. *American Psychological Journal*, 60, 209-221.
- Thornton, C., Love, J., and Meckstroth, A. (1994). *Community-level measures for assessing the status of children and families*. Presented to the annual research conference of the Association for Public Policy and Management, Chicago.
- Waits, M. J., Rex, T., and Melnick, R. (1997). *Cluster analysis: A new tool for understanding the role of the inner city in a regional economy*. Presented to the U.S. Department of Housing and Urban Development, Community Outreach Partnership Conference, Phoenix.

Wennberg, J. E., Freeman, J. L., and Culp, W. J. (1987, May 23). Are hospital services rationed in New Haven or over-utilized in Boston? *The Lancet*, 1185-1188.

White, S.B., Zipp, J. F., and McMahon, W.F. (1990). ES202: The data base for local employment analysis. *Economic Development Quarterly*.