CensusCD

Neighborhood Change Database

(NCDB)

1970 – 2000 Tract Data

Data Users' Guide
Long Form Release

Prepared by:

Peter A. Tatian

October 2003

The Urban Institute
2100 M Street, NW
Washington, DC 20037
www.urban.org

In collaboration with:

GeoLytics
www.geolytics.com

Table of Contents

1 Introduction

The Neighborhood Change Database (NCDB) contains social, demographic, economic, and housing data on census tracts in the United States for 1970, 1980, 1990, and 2000. Data in the NCDB are based on information gathered by the U.S. Bureau of the Census in its decennial censuses. The Bureau makes census tract data available to the public in both printed and machine-readable formats, but these products generally force users to focus on one tract and one characteristic at a time. By compiling data on all census tracts into one data file, the NCDB allows users to simultaneously analyze numerous (or all) tracts on a host of dimensions.

Census tracts are locally determined geographic units, typically including between 2,500 and 8,000 persons. Tracts are meant to approximate "neighborhoods" by capturing a group of residents with similar social characteristics, economic status, and housing conditions. For 1970, the NCDB contains data from 34,586 census tracts; in 1980, there are 43,221 tracts; in 1990, 61,258 tracts; and in 2000, 65,443 tracts.[1] In all four years, variables range from very general characteristics (e.g., number of persons and housing units in a tract) to detailed cross-tabulations of subgroups (e.g., race/ethnicity by educational attainment by employment status).

A powerful feature of the NCDB is its ability to match tracts across all four census years, enabling users to examine changing tract characteristics between 1970 and 2000. For many tracts, a unique identification code applies to the same physical area in all four years. But because the boundaries of many tracts change between the decennial censuses (either splitting into several tracts, merging with other tracts, or appearing for the first time), a methodology has been developed to link such tracts and their associated data to standard geographic boundaries. This is not an insignificant issue. An analysis of 1990 and 2000 census tract boundaries showed that about 49 percent of all tracts were redefined between these two census years.

While census tracts are the NCDB's basic unit of observation, each tract is also be identified according to the city, county, and metropolitan area in which it is located. Like tracts, the definitions of these areas also change from time to time, making comparisons across years difficult. The NCDB therefore allows consistent analysis of data for these larger geographies, too, permitting users to aggregate data from census tracts to broader levels of geography.

Conversely, users can append information exported from the NCDB to other data sources at the state, county, tract, city, or metropolitan area level.

This guide provides an overview of the history of the NCDB project, details about the geographic units of analysis and data fields it includes, and information to help users understand how these data were created and how they can be used effectively. (A separate Users Guide explains how to use the data access, analysis, and mapping software on the NCDB CD-ROM.)

History and Past Uses of the NCDB

The NCDB builds upon the Urban Institute's Under Class Data Base (UDB), which was created in 1989 by Isabel Sawhill and Erol Ricketts with the support of the Rockefeller Foundation. Initially, the UDB contained data for census tracts in the United States from the 1980 decennial census. It was later expanded, with further support from Rockefeller, under the supervision of Ronald Mincy and Susan Wiener to include 1970 and 1990 data.

During the 1990s, Urban Institute researchers used the UDB to analyze trends in the growth and composition of under class and concentrated poverty areas.[2] Under class areas, as measured by the Urban Institute, were census tracts that simultaneously scored high on four indicators: the proportions of female-headed families, high school dropouts, males not attached to the labor force, and welfare recipients. Concentrated poverty areas (sometimes called extreme poverty areas) were defined as tracts in which 40 percent or more of the residents live below the official poverty threshold. Urban Institute and other researchers also have used the UDB to examine changes in the geographic concentration of child poverty, regional patterns of economic distress, and the fortunes of newly arrived immigrants.[3]

In addition to being a rich data source for social science research, the UDB has provided valuable information to state and local officials, helping them identify distressed communities within their jurisdictions. For example, Virginia health officials have used the UDB to target neighborhoods in need of public health facilities. The Urban Institute has also made data from the UDB available to community-based organizations, assisting them in identifying potential clients and site locations for service delivery. For instance, the One-to-One Partnership used information from the UDB to target its mentoring programs to disadvantaged neighborhoods.

NCDB Long Form Release

The NCDB combines data from the original UDB with new information from Census 2000. As was the case with the UDB, the development of the NCDB is being supported by the Rockefeller Foundation. In addition to incorporating the Census 2000 data, the NCDB project aims to make the database easier to use and more readily available to a wider audience. To accomplish this, the Urban Institute has partnered with GeoLytics, Inc., a private firm specializing in the development of demographic and geographic data products.

Using geographic information system technology and taking advantage of GeoLytics' access to geographic boundary files for previous censuses, a methodology has been developed to "remap" earlier data to a standardized set of 2000 census tract boundaries. This methodology represents a significant improvement over the largely nongeographic methods used to remap tract boundaries in the original UDB file.

A standard set of indicators will be provided for each of the 65,443 census tracts in the United States. The NCDB Short Form Release, issued in August 2002, contained the full set of indicators for 1970, 1980, and 1990, but 2000 data only from the census "short form." These data included population counts by race and ethnicity, basic family characteristics (single-parent, elderly, families with children, etc.), and housing vacancy and tenure. The NCDB Long Form Release includes a full set of data elements based on the Census 2000 "long form."

The NCDB Long Form Release comes with all data on a single CD-ROM, produced and distributed by GeoLytics, Inc. The CD-ROM includes software for extracting the data into a variety of export file formats, creating basic graphs and descriptive statistics, and producing maps from the data. Instructions on the use of the software are in a separate *Users Guide* that accompanies the CD-ROM. The CD-ROM also contains a copy of this *Data Users' Guide* in Adobe Acrobat Portable Document Format (PDF).

NCDB Data Sources

The NCDB Long Form Release incorporates data from the following original data sources provided by the Census Bureau:

1970: Fourth Count Summary Tape for Population and Housing

1980: Summary Tape File 3A (STF3A) of the Population and Housing Count

1990: Summary Tape File 3A (STF3A)

2000: Summary File 3 (SF3) and Summary File 1 (SF1)

Additional documentation on each of these original sources can be found in the appendices of this guide.

Limitations of the NCDB

While the NCDB has powerful potential as a social science tool and information source, like any data source it has limitations that users should be aware of. As discussed in detail below, the NCDB is based on information provided by the Census Bureau. The decisions about what information to gather in the decennial census and how to tabulate the data play a critical role in determining the NCDB's format. The Census Bureau's selection of data itself tends to be stronger in certain areas (e.g., employment) than others (e.g., mobility patterns), and some topics are simply not covered (e.g., workers' wages). Census 2000, however, represents a marked improvement over previous years in both the scope of its coverage and the complexity of its cross-tabulations.

Furthermore, not every piece of information supplied by the Census Bureau has been included in the NCDB. Rather, we have selected data elements that can be made comparable from one census to the next or are of strong interest to policymakers and communities. Therefore, while the NCDB was designed for a broad audience of users, it may not satisfy everyone's data needs.

Using data from different years creates some difficulties because of changes in the way certain data are collected and tabulated. A key example, discussed further in chapter 4, is the change in the way data on race were collected for Census 2000. For many variables, we have attempted to find equivalent measures in all four years, but it is not always possible to do so. Consult the data dictionary to verify that variables are indeed consistently defined across years.

It is also important to recognize that the NCDB does not provide information on individuals directly—all data are aggregated to the census tract level. The Census Bureau aggregates data to preserve the confidentiality of individual respondents, which is guaranteed by federal law.[4] Thus, while the NCDB can be used to measure the percentage of a tract's population occupied as professionals and the tract's overall median income, it cannot directly reveal the median income of professionals. Similarly, changes in tract characteristics over time

indicate little about the fortunes of the individuals initially residing in the tract, since large numbers of people migrate into and out of tracts between the decennial censuses. For certain variables, cross-tabulations by characteristics such as race, family structure, and education provide some information about relationships between variables, but at a general level.

A final caveat concerns the nature of the NCDB's source: the decennial census. All data in the NCDB were collected at one of four points in time: 1970, 1980, 1990, or 2000. Although data from these periods are useful in performing certain dynamic analyses, doing so reveals little about the intervening periods, during which many important and interesting changes may have taken place. For some variables that move relatively slowly and steadily over time (such as housing tenure), this may not be a significant problem; but for highly variable or cyclical data (such as employment), the data from four points 10 years apart probably hide as much as they uncover—if unemployment rates were low at the end of decades but high in the interim, simply examining data from the census would be misleading.

Acknowledgments

Neither this publication nor the NCDB could have been possible without the contributions of many individuals. We acknowledge here the pioneering work of the developers of the original UDB: Ronald Mincy, Susan Wiener, Erol Ricketts, and Isabel Sawhill, the researchers who led the project at different stages; programmers Ka-Ling Chan, Mary Lee, Neal Jeffries, Gary Gerhart, Bill Marton, and George Chow; and assistants Mary Coombs and Jamie Smarr. Mitch Tobin wrote the original *UDB Users Guide*, on which this guide is based.

The directors of the current NCDB project at the Urban Institute are Tom Kingsley and Peter Tatian. Additional input, guidance, review, quality control, and data analysis were provided by Kathy Pettit, Jessica Cigna, Elizabeth Cove, Audrey Droesch, Christopher Hayes, Deborah Kaye, Alisa Wilson, Laura Harris, and Lynette Rawlings.

For GeoLytics, the NCDB development team was led by Craig Cornelius, who developed the procedures necessary to remap the 1970, 1980, and 1990 tract data to 2000 tract boundaries. Alex Vasilev was the key programmer responsible for producing the NCDB data, Natasha Vasilev developed the NCDB interface, and Katia Segre Cohen and Lynn Swartley tested the software and revised the CD-ROM *Users' Guide*.

We would also like to thank the following persons for serving as beta testers for the NCDB Short Form Release CD-ROM and providing us with many valuable comments: Michael Barndt, Cynthia Cunningham, Paul Jargowsky, and Sandra Padilla.

Finally, we are especially grateful for the generous support of the Rockefeller Foundation, which has sponsored both the development of the NCDB and a series of analytical reports based on the NCDB data to be released by the Urban Institute. For more information on these reports, visit the Neighborhood Change in Urban America web page: http://www.urban.org/nnip/ncua.

About the Urban Institute

The Urban Institute is a nonprofit, nonpartisan policy and research organization established in Washington, D.C., in 1968. Its staff investigates the social and economic problems confronting the nation and assesses government policies and programs designed to alleviate them.

The goals of the Institute are to sharpen thinking about societal problems and efforts to solve them, to improve government decisions and performance, and to increase citizen awareness of important public choices. Through work that ranges from broad conceptual studies to administrative and technical assistance, Institute researchers contribute to the store of knowledge and the analytic tools available to guide decision making in the public interest.

A multidisciplinary staff of about 400 works in nine policy centers: Education Policy, Health Policy, Income and Benefits Policy, International Activities, Justice Policy, Labor and Social Policy, Metropolitan Housing and Communities, Nonprofits and Philanthropy, and Population Studies. In addition, the *Assessing the New Federalism* project—with its own staff and mandate—operates as a cross-cutting policy center to analyze the devolution of responsibility for social programs from the federal government to the states.

To find more information about the Urban Institute and to download copies of our latest research reports, please visit our web site at http://www.urban.org.

About GeoLytics

GeoLytics is one of the leading developers of census data products in the world. Over the past six years, GeoLytics has produced and marketed a full line of "data access tools" under the label CensusCD. CensusCD products make it easy to acquire, organize, and use U.S. census demographic data, estimates, projections, and the Consumer Expenditure Survey, as well as geographic information system (GIS) data, such as TIGER cartographic files.

Over 3,200 libraries, universities, and government agencies trust GeoLytics to provide them with accurate, easy-to-use census-based demographic data. GeoLytics' customers work in many environments—in libraries; in real estate, insurance, banking, telecommunications, healthcare, media, and retail firms; in state, local, and Federal governments; and in hundreds of colleges and universities across the country.

In addition to providing demographic data packages, GeoLytics develops custom data projects. These projects have ranged from incorporating new data with customers' existing databases to creating entire systems with tools for accessing, mapping, and reporting data.

To learn more about GeoLytics and our products and services, please visit our web site at http://www.geolytics.com.

Further Support

The NCDB CD-ROM is being produced and distributed by GeoLytics. Any questions regarding purchasing copies of the CD-ROM, installation and use of the software, extraction of the data, or other technical matters should be addressed to GeoLytics:

Purchasing Products:
Email: orders@geolytics.com
Phone: 800-577-6717
Web: http://www.geolytics.com

Installation and Technical Support:
Email: support@geolytics.com
Phone: 800-577-6717
Web: http://www.geolytics.com

Questions regarding the data sources, content of the data, and potential uses may be sent to the Urban Institute by e-mail at ncdb@ui.urban.org.

Notes

[1] Not all areas of the United States were divided into census tracts in 1970 and 1980. For more on this, see chapter 4.

[2] See Ricketts and Mincy (1989), Mincy and Wiener (1993), Mincy (1993), and Galster, Mincy, and Tobin (1993).

[3] See Tobin (1993) and Zimmermann and Tobin (1995).

[4] In the 1970 and 1980 censuses, the Census Bureau also used data suppression in cases where the number of respondents was below a threshold level. See chapter 4 for more on data suppression.

2 Geography

The basic geographical unit of observation in the NCDB is the census tract. Census tracts are locally determined geographic units, ranging in size from 2,500 to 8,000 persons. Tracts are meant to approximate “neighborhoods” by capturing a group of residents with similar population characteristics, economic status, and living conditions. Tracts can be used by themselves as units of analysis or as the building blocks to create larger neighborhood areas.[1]

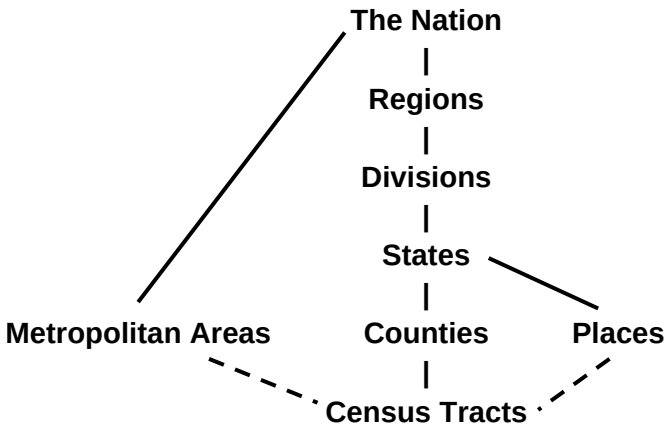
A related geographic unit also used by the census is the block number area (BNA). BNAs were established for the 1990 census to fill in the remaining areas of the country not already covered by census tracts. With the 2000 census, all BNAs have been redesignated as census tracts. This guide will use the term “census tract” or “tract” to refer to both census tracts and BNAs.[2]

Analyzing data at the neighborhood level allows users to examine the characteristics and influences of an individual's immediate surroundings. Research has demonstrated definite associations between individual behavior and neighborhood conditions and has related vulnerability to poverty to a person's community.[3] Furthermore, many social problems such as crime; substance abuse and drug trafficking; AIDS and tuberculosis; and homelessness are highly concentrated in certain neighborhoods. Any cursory glance at the United States confirms that class and race are highly spatially segregated; there are distinctly white, black, or Hispanic communities and rich, working class, or poor neighborhoods.[4]

Many proposals to ameliorate poverty, particularly in the urban environment, also stress the importance of neighborhood: enterprise zones, community-based banks, and economic development corporations all focus on the local community, its problems, and its resources.

In addition to neighborhoods or tracts, other types of geographical units are of interest to different users. States, counties, metropolitan areas, and other types of areas can be useful for a variety of analyses. This chapter describes the various types of geography included in the NCDB and discusses the issues involved in working with historical census data at these different levels.

Figure 2-1: Hierarchy of Census Geography in the NCDB



Geographic Identifiers in the NCDB

Geography is organized hierarchically in both the census, and the NCDB. Figure 2-1 provides a basic organization of the census geographical hierarchy as implemented in the NCDB. Not all these levels are available in each of the four census decades. It is important to note as well that "the nation" in the 1970 and 1980 NCDB data does not represent the entire United States. Data from these two decades are only for "tracted" land, which includes all metropolitan counties and a very limited number of nonmetropolitan areas. Thus, large expanses of sparsely populated rural land are not part of the NCDB for 1970 and 1980. For 1970, the NCDB includes 148,456,474 persons, or 73 percent of the 203,302,031 Americans actually enumerated in the 1970 census; in 1980, the NCDB covers 181,171,224 persons, or 80 percent of the 226,545,805 total. In 1990 and 2000, however, all land in the United States is covered, either as census tracts or as BNAs, and therefore the NCDB includes all persons counted by the census.

Each observation of tract data in the NCDB has a series of geographic identifier variables that associates the tract with various other geographic levels. These identifier variables often contain numeric codes that correspond to a particular unit of geography. For example, the variable STATECD has the value “39” for all tracts located in the state of Ohio. Note that, although the codes may consist of numbers, the geographic identifier variables are

actually all stored as character values in the NCDB. Smaller numbers are “zero padded” so that they have their full length. For instance, the STATECD value for Alabama is “01”. If the NCDB data are exported to an external file, maintaining these identifiers as character variables is important to be able to join these data with standard geographic information system (GIS) spatial data files.

Using these geographic identifiers, the NCDB tract-level data can be combined or “aggregated” from one level to another if there is a line connecting those levels in figure 2-1.[5] Because of the way these levels are constructed, however, not all of the aggregations will match perfectly with census published numbers for these levels. These cases are indicated by the dashed lines in figure 2-1 and are noted in the sections describing each level below.

Regions and Divisions

The Census Bureau divides the nation into four regions: Northeast, South, Midwest, and West.[6] These four regions are further subdivided into nine divisions: New England and Middle Atlantic (in the Northeast); West South Central, East South Central, and South Atlantic (in the South); West North Central and East North Central (in the Midwest); and Mountain and Pacific (in the West). The 50 states and the District of Columbia are the components of regions and divisions and are allocated as shown in figure 2-2.

Figure 2-2: Census Regions and Divisions

Northeast Region (1)

New England Division (1): Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut

Middle Atlantic Division (2): New York, New Jersey, Pennsylvania

Midwest Region (2)

East North Central Division (3): Ohio, Indiana, Illinois, Michigan, Wisconsin

West North Central Division (4): Minnesota, Iowa, Missouri, North Dakota, South Dakota, Nebraska, Kansas

South Region (3)

South Atlantic Division (5): Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida

East South Central Division (6): Kentucky, Tennessee, Alabama, Mississippi

West South Central Division (7): Arkansas, Louisiana, Oklahoma, Texas

West Region (4)

Mountain Division (8): Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, Nevada

Pacific Division (9): Washington, Oregon, California, Alaska, Hawaii

States in the same division and, to a lesser extent, in the same region tend to have similar economic bases; racial/ethnic composition; climate and geography; and political preferences. Nevertheless, this partition of the nation is very general, and much variation exists within divisions and regions.

In the NCDB, the variables REGION and DIVIS identify a tract's census region and division. Each is identified by a single digit numeric code, as shown in figure 2-2. The classification of regions and divisions is available for all four census decades and does not change across years. Tract-level data can be aggregated to regions and divisions with complete accuracy.

States

The NCDB assigns a set of unique state codes to each of the 50 states and the District of Columbia. The variable STATECD indicates a two-digit Federal Information Processing Standards (FIPS) code for each state, from "01" to "56". The variable STATE contains the same information.[7] FIPS codes are based on the alphabetical order of state names. The variable STATECE contains the two-digit census state code, which is different from the FIPS code and ranges from "06" to "95". The census state codes are ordered by region and division.[8] Finally, the variable STUSAB contains the two-letter U.S. Postal Service state name abbreviation.

A complete set of state codes and names can be found in appendix A. The classification of states is available in all four census decades and does not change across years. Tract-level data can be aggregated to states with complete accuracy.[9]

Counties

Counties are the basic political subdivision of most states, with a few exceptions. Louisiana is divided into "parishes" and Alaska into "boroughs." These subdivisions are treated as county equivalents in the census and in the NCDB. In four states (Maryland, Missouri, Nevada, and Virginia), some cities are independent of any county; all of these "independent cities" are considered counties in the census and in the NCDB. The District of Columbia has no county subdivisions and is thus treated as a single county.

For the 1990 and 2000 censuses, there were 3,141 counties (or county equivalents) in the United States. Tracts from all of these counties are included in the NCDB for both decades. For 1970 and 1980, only 615 and 911 counties, respectively, are represented because the NCDB does not contain information for the "untracted" areas in these years. Most of the omitted counties were in nonmetropolitan areas.

Each county is assigned a three-digit FIPS code, which appears in the NCDB as variables COUNCD and COUNTY. County codes are unique to each state and are ordered according to the alphabetical order of the county names (independent cities, however, appear after the listing of counties). It is important to note that county codes repeat across states—every state has its own "001" county. Therefore, to access information about tracts in a given county, one must supply both the state code and the county code. To make this easier for

NCDB users, the variable UCOUNTY contains a unique five-digit county code consisting of the two-digit FIPS state code and the county code.

Appendix B lists the FIPS codes for all counties in the United States. The classification of counties is available in all four census decades, with the caveat noted above for 1970 and 1980. County boundaries can change across years, although this happens rather rarely. To compare county-level data for consistent geographic boundaries over time, use the 1970, 1980, and 1990 data remapped to 2000 tract boundaries, which are based on 2000 county definitions. Within any given year, tract-level data can be aggregated to counties with complete accuracy.

Census Tracts

Census tracts are the NCDB's basic unit of observation. They are small, relatively permanent divisions drawn by local census statistical areas committees in accordance with Census Bureau guidelines. When their boundaries are first determined, census tracts are meant to be "homogeneous with respect to population characteristics, economic status, and living conditions." Census tracts do not cross county or state boundaries but may cross other types of jurisdictions. Except in New England, census tracts do not cross metropolitan area boundaries.

Census tracts are classified by a two-part identification number, which consists of a four-digit code and, in some cases, a two-digit suffix. In Census Bureau publications, the code and suffix are often printed with a decimal point separating them, as in "1234.01". In the NCDB, the first four-digit code number is the variable TRCTCD1 (in the example, "1234"), while the two-digit suffix is TRCTCD2 (in this example, "01"). The decimal point is not included. If the two-digit suffix is “00”, TRCTCD2 is blank.[10]

A third tract variable, TRACT*yy*, contains the four-digit code and two-digit suffix together, without the decimal point (e.g., “123401”). The variable name includes the two-digit year of the tract definition (e.g., TRACT90), to avoid confusing tracts from different decades.

Census tract numbers are unique only within counties. Thus, to identify a specific tract out of all those in the United States, one must provide the state and county codes in addition to the tract identification number. The variable GEO*yyyy* contains an 11-digit code that includes all of this information and, therefore, uniquely identifies each tract (e.g., “01003123401”). The variable name includes the four-digit year (*yyyy*) of the tract definition (e.g., GEO1990).

Some redrawing of census tracts occurs between each decennial census. Boundaries are drawn with the intent of being maintained over time in order to facilitate comparisons, but physical developments (such as new housing complexes) and extreme population changes force some tracts to be added, split, or merged with other tracts. To enable users to compare tracts over time, the NCDB has remapped 1970, 1980, and 1990 tract data to Census 2000 tract boundaries. Data for the earlier decades are either available in their original tract boundaries or are remapped. For more information on the method used to remap the data and on the comparability of tract boundaries between years, please see chapter 4.

Metropolitan Areas

Metropolitan areas (also called metro areas) are a complicated issue. Some of the complication arises from a change in classification initiated between the 1980 and 1990 censuses, while some arises from the difficult task of operationalizing the notion of a metropolis.

Metropolitan area definitions are actually determined by the Office of Management and Budget (OMB), which follows official standards created by the interagency Federal Executive Committee on Metropolitan Areas. At the heart of every metro area is at least one "population nucleus." The basic requirement for a region to be designated a metro area is the presence of a city with at least 50,000 residents or a census-defined "urbanized area" in which the constituent counties have at least 100,000 residents (or 75,000 in New England). The determination of a metro area attempts to capture this center and the surrounding areas that are economically or socially connected to it. Outside of the New England states, counties are the building blocks for metro areas; within New England, towns and cities are used instead. If a county (or, in New England, a city or town) meets certain standards for economic or social integration with the central county or place, then the entire county (or city or town) is included in the metro area.[11]

For the 1990 and 2000 censuses, there were three types of metro area designations: Metropolitan Statistical Areas (MSAs), Consolidated Metropolitan Statistical Areas (CMSAs), and Primary Metropolitan Statistical Areas (PMSAs). Metropolitan areas with less than one million persons are simply designated as MSAs—they are usually not closely tied to other MSAs and are generally surrounded by nonmetropolitan counties. If a metro area has over 1 million residents, it may be designated as a CMSA and divided into PMSAs. A PMSA is typically a "large urbanized county or cluster of counties that demonstrates very strong internal economic

and social links, in addition to close ties to other portions of the larger area." At the time of the 2000 census, there were 258 MSAs, 18 CMSAs, and 73 PMSAs.

To illustrate this classification system, consider the New York metropolitan area. There is one "New York-Northern New Jersey-Long Island, NY-NJ-CT-PA CMSA," which includes the 21 million people living in New York City and its surrounding metropolis. This CMSA is divided into 15 PMSAs, such as Bergen-Passaic, NJ; Bridgeport, CT; and Nassau-Suffolk, NY. In contrast, the Buffalo-Niagara Falls, NY MSA includes the 1.2 million people living in Erie and Niagara counties, but it is not considered to be directly linked to any other metro area.

Before June 1984, the equivalent of MSAs were called Standard Metropolitan Statistical Areas (SMSAs), and CMSAs were known as Standard Consolidated Statistical Areas (SCSAs)—there was no equivalent to PMSAs. These designations are used in the 1970 and 1980 census data.

In addition to the standard metro areas, in New England an alternative set of metro areas has been defined, which consist of aggregations of counties rather than of cities and towns. This allows the use of metro area definitions in New England that are consistent with those in the rest of the country. These alternative metro areas are called New England County Metropolitan Areas (NECMAs).[12] In 2000, there were 12 NECMAs.

Metro areas are identified by a series of unique codes. CMSAs have both two-digit and four-digit identification codes; all other metro area designations have four-digit codes. The codes are unique across the entire United States and also across types. So, no PMSA will have the same identification number as any CMSA, MSA, or NECMA, nor will any other type of metro area have the same code as one of a different type.

In the 1970 and 1980 NCDB data, there is just one geographic identifier for metro areas: SMSA*yy*. This variable contains the four-digit code for an SMSA. The variable name includes the two-digit year (SMSA70 or SMSA80) to emphasize that metro definitions and identification codes may change over time.

In 1990 and 2000, several variables are included to identify the different types of metro area designations adopted in 1984. Variable CMSA*yy* contains the two-digit code identifying CMSAs. The variable name includes the two-digit year (*yy*) corresponding to the metro area definition being used.[13] Variable MSACMA*yy* is the four-digit MSA code, if the tract is in an

MSA, or the four-digit CMSA code, if the tract is in a CMSA/PMSA. Variable MSAPMA*yy* is the four-digit MSA code, if the tract is in an MSA, or the four-digit PMSA code, if the tract is in a CMSA/PMSA. Finally, the variable PMSA*yy* contains the four-digit PMSA code, if the tract is in a CMSA/PMSA.

To identify NECMAs, the variable NECMA*yy* contains the four-digit NECMA code.

If the tract is not in a metro area or not in one of the appropriate type, then the identification variable will have the value "9999" (or "99" for the two-digit CMSA codes).

Appendix C lists the codes for the metro areas in the United States. Metro area definitions are updated periodically, usually more often than once every 10 years. Metro areas can expand or shrink, depending on changes in demographic and economic patterns over time. In most cases, metro areas keep the same codes from 1990 to 2000, but some changes or redefinitions do occur. It is therefore very important to be clear on which metro definition is being used, especially when working with historical data.

Tract-level data can be aggregated to MSAs, CMSAs, and PMSAs with complete accuracy outside the New England states. Within New England, tract data can be aggregated with complete accuracy only to NECMAs. Nevertheless, the aggregation error for metro areas in New England is quite small (between –0.5 and 0.9 percent of the total population). See appendix D for a summary of the tract aggregation error for New England metro areas.

Central Cities

All metro areas include a central city area, which constitutes the nucleus of the region. The designated "central city" can actually include multiple cities or urban counties that constitute this nucleus. The relationship between the central city and the rest of the metro area is often of great interest to analysts studying urban problems.[14] Considerable research has focused on inner-city poverty, the exodus of whites and middle-class blacks from the central city, and the imbalances of tax bases throughout the metro area. In 1990 and 2000, the variables PCMACC*yy* and PCNECC*yy* indicate the percentage of the tract population who were living in the central city or cities of a metro area.

Places

"Places" include both incorporated cities and towns and census-designated places (CDPs). CDPs are the statistical equivalents to incorporated areas that "comprise densely settled concentrations of population that are identifiable by name, but are not legally incorporated places." The name assigned to a place identifies whether it is a city, town, or CDP. There were 25,150 places identified in the 2000 census.

The NCDB contains place identifiers for the 1980, 1990, and 2000 data. The variable PLACE*yy* contains the five-digit (four-digit in 1980) FIPS place code, which is unique within a state and assigned according to the alphabetical order of a state's place names. A complete list of current place codes can be found on the FIPS web site at http://www.itl.nist.gov/fipspubs/55new/nav-top-fr.htm. To identify a place uniquely, it is necessary to combine the state code with the FIPS place code. For the convenience of users, the NCDB provides the variable UPLACE*yy*, which combines the state and place codes into a single seven-digit value. Tracts that are not located within an identified place are given a PLACE*yy* code of all 9's and a UPLACE*yy* code of the two-digit state code and the remainder 9's.

For the 1980 data, the variable PLCDSC80 contains a single character code indicating the type of place where the tract is located. The list of these codes can be found in figure 2-3.

Figure 2-3: Place Description Codes, 1980

Code (plcdsc80)	Place Description
1	Incorporated central city of SMSA not urbanized area (UA)
2	Incorporated central city of UA not SMSA
3	Incorporated central city of SMSA and UA
4	Other incorporated place
9	Not place; part of MCD/CCD
A	Census designated place (CDP), central city of UA not SMSA
B	CDP, central city of SMSA and UA
C	CDP in UA with central city of 50,000 or more
E	CDP coextensive with MCD or county
F	CDP of 1,000 or more, not in UA or CDP in UA with central city of 50,000 or less
G	CDP in Hawaii or outlying areas

Place definitions can change at any time and so are difficult to track. Since tracts can cross place boundaries, tract-level data cannot be aggregated to place level with complete accuracy. For the NCDB, we assigned a place identifier to each tract based on where the largest percentage of the population lived at the time of the census. It is best to compare population or other aggregated totals with published Census Bureau numbers for particular places to be aware of any serious errors that may result. See appendix D for a summary of tract aggregation areas for the principal cities in the 100 largest metro areas.

Urban Areas

Whether or not a tract is within a metro area does not indicate the extent to which a tract is urban or rural. The "urban/rural" distinction used by the Census Bureau is separate from that of the metro area determination. To give users a measure of urbanization, the NCDB variable PCURBN00 indicates the percentage of the tract's 2000 population who were living in a census-designated urban area.

Other Geographic Units

A variety of other geographical identifiers are included with the 2000 data. They are listed in figure 2-4. Most do cross tract boundaries and therefore will not aggregate with complete accuracy from the tract level. Users should compare their own tabulations with published census sources to determine the level of error.

Figure 2-3: Other Geographic Identifiers for 2000

Variable	Description
COUSUB00	County Subdivision (FIPS) MCD/CCD
AIANHH00	Am. Indian area/AK Native Area/HI Home Land
CD106	Congressional District (106th)
SLDU00	State Legislative District (Upper Chamber)
SLDL00	State Legislative District (Lower Chamber)
SDELM00	School District (Elementary)
SDSEC00	School District (Secondary)
SDUNI00	School District (Unified)
VTD00	Voting District
ZCTA500	ZIP Code Tabulation Area (5 digit)
AREAKEY	Tract Number
AREALAND	Land Area (sq. meters)
AREALANM	Land Area (sq. miles)
AREAWATR	Water Area (sq. meters)
AREAWATM	Water Area (sq. miles)
INTPTLAT	Internal Point (Latitude)
INTPTLON	Internal Point (Longitude)

Notes

[1] For more on the definitions on tracts and other census geographies, see the summary file documentation for the appropriate year, or U.S. Census Bureau (1994).

[2] For more on how tracts are defined, see U.S. Census Bureau (2001), appendix A.

[3] See Crane (1991), Clark (1992), Ellen and Turner (1997), and Wilson (1987).

[4] See Massey and Denton (1993).

[5] Any aggregations of NCDB data must be done by exporting data from the NCDB and using an external software product (such as SAS or SPSS) to perform the aggregation.

[6] Before 1984, the Midwest region was referred to as North Central.

[7] Two versions of the same variable are provided for compatibility purposes. STATECD is consistent with the previous UDB data; STATE matches the naming scheme used by the Census Bureau.

[8] While the Census Bureau treats the outlying areas—American Samoa, Guam, the Northern Mariana Islands, Palau, Puerto Rico, and the Virgin Islands of the United States—as equivalent to states, these areas are not included in the NCDB for any year.

[9] For 1970 and 1980, tract data aggregated to the state level will not necessarily match the state totals reported by the census because not all of the country was covered by census tracts in those years.

[10] The two-digit suffix can sometimes provide useful information about the tract. Tracts with a suffix of "99" are populated entirely by persons aboard one or more civilian or military ships. Suffixes ranging from "80" through "98" generally identify tracts that were revised or created for the 1990 census, some of which have little or no land mass or population.

[11] For more information, see FIPS (1995) and OMB (1990).

[12] It should be noted that NECMAs are in fact used relatively rarely by people doing metropolitan area analysis in preference to the MSA/CMSA/PMSA definitions.

[13] Metro definitions are not updated every 10 years but can be changed at various times throughout the decade. Therefore, it is very important to be clear about which metro definition is being used, especially when analyzing historical data. The metro definition for 1990 census data is from 1990, while the one for 2000 data is from 1999.

[14] Many researchers refer to the part of the metro area outside the central city area as the “suburbs.” It is important to note, however, that neither the Census Bureau nor the Office of Management and Budget has an official definition for suburban areas.

3 Data Dictionary

Appendix E of this guide is the "data dictionary" for the NCDB. It lists all of the data fields or "variables" available for 1970, 1980, 1990, and 2000; gives a brief description of their definitions; and indicates the census source data used to construct them. The data dictionary is an important resource for users as it tells them exactly how the NCDB variables were formed from the original census tabulations. Users should always consult the data dictionary if they have any questions about a variable's precise definition or its comparability across census years.

Classification of Variables

In accordance with Census Bureau conventions, NCDB variables are separated into two general groups: population and housing. Population variables are further classified into nine categories:

General Population Characteristics
Family Structure/Marriage
Mobility/Transportation
Education
Employment/Labor Market
Poverty/Public Assistance
Income and Earnings
Age Distribution
Language Ability

While housing variables are grouped into four categories:

Housing Tenure/Occupancy
Housing Characteristics/Utilities
Housing Costs/Affordability - Owners
Housing Costs/Affordability - Renters

Additionally, there is a category for geographic identifier variables.

To make it easier to find data of interest, variables in the data dictionary are grouped according to their category. Users should be aware that some closely related variables may actually be found under different categories in the data dictionary. Some variables in the Age

Distribution category are also found elsewhere. These variables are typically denominators of ratios that specify a certain age bracket.

When selecting variables with the NCDB access software on the CD-ROM, users can also search for variables according to keywords, which may make it possible to find related variables more quickly than searching through the data dictionary. For example, one could search for all of the variables associated with the key words “black” and “renters” or with “families” and “poverty.”

Sample Entry

A sample data dictionary entry is shown in figure 3-1. Each entry briefly describes a set of related variables across one or more census years and provides their data sources. Because of changes in the way decennial census data are collected and tabulated, not all variables will be available for all four decades. Variables in the same entry are meant to be comparable across the years, although for some variables the match is not perfect. Different variable descriptions can alert the user to obvious changes in variable definitions, but it is recommended that users refer to the original census documentation to check on comparability between years (see below for more on data sources).

Figure 3-1: Sample Data Dictionary Entry

AVEMER*y*D ❶ ❷ 1970, 1980, 1990, 2000

Households with wage and salary income last year ❸

Variable	*Year*	*Source/Calculation*
AVEMER7D ❹	1970 ❺	Table P80(1):1,7 ❻
		Suppr. flags: PFLG18 ❼
AVEMER8D	1980	Table 71:2
		Suppr. flags: AFLAG30
AVEMER9D	1990	Table P90:1
AVEMER0D	2000	Table P59:2

Notes: 1970: Families and unrelated individuals 14+ years old. ❽

The first two items in a data dictionary entry are located above the solid horizontal line. These are the general variable name ❶ and the list of years for which the variable is available

❷. In the sample in figure 3-1, the general variable name is "**AVEMERyD**" and this variable is available in all four years. The general description of this variable is given in ❸. Below the description, is the complete set of variable names ❹, listed by year ❺. Along with the name and year, the data dictionary entry includes information about how the variable was created ❻— either from the original census tabulations or as a calculation of other NCDB variables. In the sample entry, the variables were all created directly from census tabulations, as specified. Certain variables may be derived from other NCDB variables, such as the proportion of white persons in a tract in 2000, which is calculated as "SHRWHT0N / TRCTPOP0".

Two other items appear in this entry, the data suppression flags that apply to the source information for this variable in 1970 and 1980 ❼, and a note providing further information about the variable ❽. In this case, the note points out a definitional difference of this field in the 1970 census.

The remainder of this chapter contains more information about the naming of NCDB variables and how to understand the census source descriptions.

Variable Names

Variable names were created with a few basic rules in mind. All variable names are eight characters or less to make them compatible with almost all database and data analysis software. Variable names include only the uppercase letters (A-Z), numbers (0-9), and the underscore character ("_"). This obviously limits how descriptive the name itself can be, which is why it is important to consult the description provided in the data dictionary.

The basic structure of a variable name is shown here, with the example BLKPR7DA:[1]

BLKPR	**7**	**D**	**A**
Base	*Year*	*Num/Den suffix*	*Opt. suffix*

The "base" part of the name consists of up to seven letters and numbers. This is followed by a single-digit number that indicates the census year of the data. Data for 1970 have the number "7" included in their name, while data for 1980 have an "8", for 1990 a "9", and for 2000 a "0". Thus, for the base name "TRCTPOP", we have the variables TRCTPOP7, TRCTPOP8, TRCTPOP9, and TRCTPOP0, which are the total population in 1970, 1980, 1990, and 2000, respectively.

Variables used as numerators in proportions or ratios may also have the numerator suffix "N" after the year digit in their names, while variables used as denominators in ratios may have the suffix "D". In these cases, the calculated proportion or ratio may have the same (or similar) base name without the numerator/denominator suffix. For example, WELFAR9N is the number of households in a tract receiving public assistance in 1989, while WELFAR9D is the total number of households in the tract. Dividing these two (that is, WELFAR9N / WELFAR9D) gives us the variable WELFARE9—the proportion of households receiving public assistance in 1989.

Finally, a few 1970 variables have an optional single-letter suffix tacked on to the end of the variable name to distinguish it from another version of the same data. For example, the variable BLKPR7DA is an alternative denominator for the 1970 black poverty rate that excludes persons living in group quarters (that is, mental hospitals, nursing homes, military barracks, college dormitories, and other institutions.)

Users will also note that similar abbreviations are used in variable names for characteristics that are common in the NCDB data. For example, "BLK" or "B" is typically used to abbreviate blacks, "WHT" or "W" for whites, and "HSP" or "H" for Hispanics.

Source Tables and Cells

The "Source/Calculation" column in the data dictionary entry lists either the tables and cells used to create the variable from the source census tabulations or the formula used to compute the variable from other NCDB data. For variables created directly from census tabulations, it can be very important to refer to the original census definitions to understand exactly what the data represent. The complete set of source tabulation matrices for the 1970–2000 decennial census data are included in appendices F through I of this Data Users Guide.

Census tabulations change from year to year both in the numbers and types of tabulations and in the way the matrices are numbered and identified. Nevertheless, there are some consistencies across years. A census summary tabulation file consists of a series of tables, which contain a set of data elements (“cells”) relating to a particular subject. Tables are listed numerically and have descriptive titles. In 1970, 1990, and 2000, tables for housing variables are numbered separately from those for population variables, while in 1980 the housing tables follow the population tables with no breaks in numerical order. For 1970, there

were 127 population tables and 200 housing tables; for 1980, there were 150 tables in total; for 1990, population variables are listed in tables P1 through P170, while housing variables are listed in tables H1 through H92; and for the 2000 data used to create the Long Form Release, there were 484 population tables and 329 housing tables. In all years, the table title includes any characteristics used to cross-tabulate the data; for example, “Employed Females by Occupation” would list the breakdown of all employed females into assorted occupational categories.

A table's "universe" is listed below its title. The universe describes the unit of observation for the table. For example, the universe for a table might be all persons, all persons above or below a certain age (such as adults or children), all households, or all renter-occupied housing units.

Cells are the building blocks of tables. If all the values from all of the unduplicated cell counts in a table are added up, the sum is equal to the universe for the table. Prior to 2000, this is the same as adding up all of the cells in the table. The Census 2000 tabulations, however, always include the universe total as the first cell, and then may also include subtotals for cross-tabulation categories (such as males and females). Therefore, one needs to eliminate these duplicate counts to be able to add the individual cells to the universe total.

In 1970, the number of cells in a table is listed under the "No. of Data Items" column; in 1980 and 2000, the number of cells is within square brackets at the end of the title line; and in 1990, the number is listed under the "Total Number of Data Cells" column. Note that some tables have only one cell. Each cell in a table is presented as a separate line in the table matrices. Prior to 2000, lines ending in a colon (":") are *not* cells, but merely titles for cross-tabulation categories. The actual cells are the indented lines below these entries that do not end with colons. For Census 2000, however, the lines ending with a colon *do* represent actual data cells in the tabulations.

To simplify presentation, "repeat" statements are sometimes included in census table matrices. Typically, the same categorization will be duplicated for, say, several racial groups, and in order to avoid repeating this categorization, the tables will simply instruct the user to repeat the layout for each of the racial groups.

In 1970, tables presenting data that are cross-tabulated by race/ethnicity are different than similar tables in 1980 and 1990. These tables in 1970 are divided into four segments,

meaning that within each table, there is a separate record for 1) all individuals; 2) whites; 3) blacks; and 4) Spanish Americans (termed "persons of Hispanic origin" in later censuses). In the data dictionary, if no number in parentheses follows the table number, the data are for all individuals. A "(2)" indicates the count is for whites only; a "(3)" indicates the count is for blacks; and a "(4)" indicates the count is for Spanish Americans. For example, TRCTPOP7 (total population) is calculated using table 17, while SHRWHT7N (white population) uses table 17(2), SHRBLK7N (black population) uses table 17(3), and SHRHSP7N (Hispanic population) uses table 17(4).

The sample table from the 2000 summary file tabulation in figure 3-2 illustrates some of the concepts discussed above.[2] The sample is for population table “P22,” which tabulates the number of households by whether they have someone 60 years or older living there, by the number of people living in the household, and by the household type (family/nonfamily). The universe for this table is all households in the United States.

Figure 3-2: Sample Census Table Matrix

Table	Cell	Description
P22.		HOUSEHOLDS BY PRESENCE OF PEOPLE 60 YEARS AND OVER, HOUSEHOLD SIZE, AND HOUSEHOLD TYPE [11]
		Universe: Households
	1	Total:
	2	Households with one or more people 60 years and over:
	3	1-person household
	4	2-or-more person household:
	5	Family households
	6	Nonfamily households
	7	Households with no people 60 years and over:
	8	1-person household
	9	2-or-more person household:
	10	Family households
	11	Nonfamily households

This table has a total of 11 cells (as indicated in the square brackets at the end of the table title). The first cell is the “Total” number of households—the universe for the table. The second cell is the number of households with someone age 60 or older living in them. Cells 3 and 4 are indented under this category, so they refer only to households having someone age

60 or older. Cell 3 is the total number of one-person households (i.e., the number of people 60 or over living alone), while cell 4 is the number of such households containing two or more persons. This latter group is further broken down in cells 5 and 6, which contain the number of family and nonfamily households, respectively, for two- or more person households including at least one person age 60 or older. Cells 7–11 repeat the breakdowns in cells 2 through 6, but for households *without* someone age 60 or older.

As noted earlier, the data dictionary in appendix E indicates the tables and cells used to create the NCDB variables from the census source tabulations. In these specifications, the table number is listed first, followed by a colon and the cell number or numbers from that table. For example, the specification "P19:5" refers to table P15, cell 5. Sometimes several cells in a table are added (or subtracted) or several tables are used. A dash between two cell numbers following a table specification indicates that all cells between and including those points were summed. For example, "P74:4-7" is equivalent to (P74:4) + (P74:5) + (P74:6) + (P74:7). In a few cases where the same cell numbers from several related tables were used, the included tables are separated by commas. For example, 14A,C,E,G,I:11 is equivalent to (14A:11) + (14C:11) + (14E:11) + (14G:11) + (14I:11).

Notes

[1] Note that geographic identifiers and suppression flags have different naming conventions.

[2] This sample is for the SF1 tabulations.

4 Special Issues

This chapter discusses a number of special issues relating to the development and use of the NCDB data. It is necessarily more technical than some of the preceding material. The sections below cover issues relating to the geographic comparability of census data and the matching of tracts across census years, merging other data sources with the NCDB, data suppression, the census undercount, inaccurate responses, the 1990 homeless population count, bridging race data between 2000 and previous censuses, and changes in the determination of Hispanic origin.

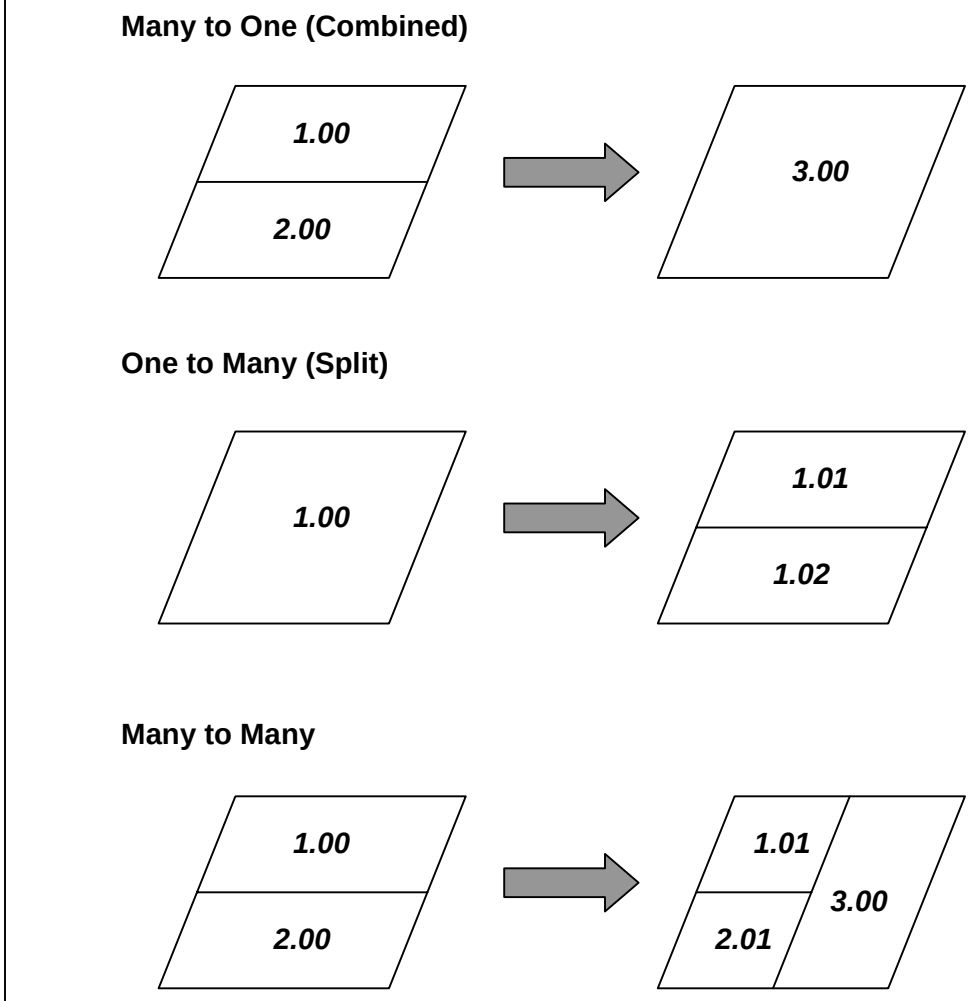
Geographic Comparability and Matching Tracts Across Census Years

One of the most valuable features of the NCDB is its ability to match tracts across census years. For many tracts, the same identification code applies to the same physical space in the 1970, 1980, 1990, and 2000 censuses—that is, these tracts have not changed boundaries between the decennial censuses. Many other tracts, however, have redefined boundaries, usually due to changes in their population. A tract that loses a significant portion of its residents will often be merged with surrounding tracts, thus altering that tract and any tracts with which it is combined. Tracts that experience rapid population growth will frequently be divided into a set of smaller tracts. Often in these cases, the four-digit tract code (TRCTCD1) is retained, while new two-digit suffixes (TRCTCD2), such as "01", "02", and "03", are added. For each new census, a few tracts are completely eliminated (if, for example, an entire area is razed), while new tracts are added to accommodate new residential areas and, between 1980 and 1990, because of the expansion of tract coverage to the entire nation.

Figure 4-1 illustrates the three main types of tract changes that occur between censuses. The first type of change is when a two or more tracts from one census year are combined to form a single tract in a subsequent census. We refer to this as a “many to one” change. The second type is when a single tract splits into two or more tracts for a subsequent census. This is referred to as a “one to many” change. The third type of change occurs when two or more tracts are reconfigured into two or more different tracts, which we call a “many to many” change. In the example of a “many to many” change in figure 4-1, the two tracts

numbered “1.00” and “2.00” are redrawn into three new tracts: “1.01”, “2.01”, and “3.00”. (A fourth type of change, not shown in the figure, occurs when a tract does not change its boundaries but is “renamed” with a different ID. We refer to this as a “one to one” change.)

Figure 4-1: Types of Census Tract Changes



These changes are not insignificant. Based on analysis of geographic data for census tracts in different years, we have determined that 49 percent of all 2000 census tracts experienced boundary changes since the 1990 census. Most of these changes are the “many to many” type (38 percent), followed by “one to many” (9 percent), and “many to one” (2 percent). As well as being the more common types of change, the “many to many” and the “one to many”

changes are the most difficult to deal with. If two or more tracts in 1990 simply were combined into a single tract in 2000 (“many to one”), then a user only needs to add together the 1990 data for these tracts to obtain the correct totals for the 2000 tract. If, however, the tract splits into one or more pieces between 1990 and 2000, then the user must know the relative proportion of the population living in the different pieces making up the 2000 tract.

The actual remapping procedure for converting data from 1970, 1980, and 1990 tracts to 2000 tract boundaries is quite complicated. Those wishing more a technical explanation of this task should consult appendix J. The basic procedure was to use geographic information system (GIS) software to overlay the boundaries of 2000 tracts with those of an earlier year. This allowed us to identify how tract boundaries had changed between censuses. We then used 1990 block data to determine the proportion of persons in each earlier tract that went into making up the new 2000 tract. For example, if a 1990 tract split into two tracts for 2000, the population may not have been divided evenly. Our method allows us to determine the exact weight to allocate to each portion.[1]

These population weights were then applied to the various 1970, 1980, and 1990 tract-level NCDB variables to convert them to 2000 tract boundaries.[2] The population weights were used to convert all variables based on counts of persons, households, and housing units, all counts based on subpopulations (such as black persons or elderly households), and all aggregate data (such as aggregate household income). Proportions (such as the proportion of Hispanic persons) were remapped by first converting the respective numerator and denominator values (Hispanic persons and total persons, respectively) and then recalculating the proportion.

The 1970, 1980, and 1990 NCDB data are available in two versions. One version is based on tract boundaries as drawn in each individual census year, that is, 1970 tract boundaries for 1970 data, 1980 tract boundaries for 1980 data, and 1990 tract boundaries for 1990 data. This is the standard format used for analyzing tract characteristics in a single year. When accessing the data through the NCDB CD-ROM, this version of the data is obtained by selecting a single year from the “Year” menu. Note that, in this case, only one year can be accessed at a time, and separate extracts must be performed to get data for more than one year.

The second version of the 1970, 1980, and 1990 data are these variables remapped or “normalized” to 2000 census tract boundaries. This version is used to match tracts and compare

their characteristics over time. The remapped version of the data is obtained by selecting "All years normalized to 2000" from the "Year" menu on the NCDB CD-ROM. One can then select variables for any of the four census years from the "Count" selection dialog. Any extract files or maps created in this manner will be normalized to 2000 tract boundaries.

Of course, there is just one version of the 2000 NCDB data, which is available only in 2000 tract boundaries. These data can be accessed on the CD-ROM through either of the methods described above, that is, selecting the year "2000" from the "Year" menu, or selecting "All years normalized to 2000" and choosing 2000 variables from the "Count" selection dialog.

Coverage for 1970 and 1980

It should be noted again that, since the source data for the 1970 and 1980 NCDB variables are the original tract-level tabulations provided by the census, which did not cover the entire United States, not all 2000 tracts will have data available for these earlier years. This may not be completely obvious from examining the data, since only ***part*** of a 2000 tract may have been covered by census tracts in 1970 or 1980. Therefore, some data might be available for the 2000 tract, but these data may not represent the entire tract area or population.

Two indicator variables are available to allow users to identify these situations. PCTCOV70 and PCTCOV80 are available with the remapped data and indicate the percentage of 2000 census blocks that were covered by 1970 and 1980 tracts, respectively. If the percentage is 100, then one knows that the 1970 or 1980 data are complete for that tract. If the percentage is less than 100, then data are unavailable for some part of the 2000 tract. When using 1970 or 1980 data, users may wish to exclude tracts that are less than 100 percent covered or those that have coverages less than some threshold percentage.

Tract Change Flag Variables

It may be important for some users to know which tracts have undergone changes between censuses and which have remained the same. To allow users to identify these tracts, three tract change flag variables are available with the remapped NCDB data. These three variables, TCH70_00, TCH80_00, and TCH90_00, indicate the extent to which 2000 tracts have changed between 1970, 1980, and 1990, respectively. These variables contain a single-digit numeric code denoting the type of tract change (if any). In addition to the three types of tract changes ("many to one," "one to many," and "many to many"), there are codes indicating

whether a tract was renamed (“one to one”), was in a nontracted area in 1970 or 1980, or did not change at all between censuses.

Figure 4-2 lists the codes for the tract change variables and summarizes the extent of the tract changes between the three earlier census years and 2000.

Figure 4-2: Summary of Census Tract Changes

	1970 to 2000		1980 to 2000		1990 to 2000	
Tract Change Status	**No. Tracts**	**Pct.**	**No. Tracts**	**Pct.**	**No. Tracts**	**Pct.**
Total 2000 Tracts	65,232	100.00	65,232	100.00	65,232	100.00
0) No change	13,841	21.22	18,891	28.96	32,129	49.25
1) 1 to 1 (renamed)	936	1.43	1,226	1.88	1,076	1.65
2) Many to 1 (combined)	250	0.38	397	0.61	976	1.50
3) 1 to many (split)	7,282	11.16	6,920	10.61	6,106	9.36
4) Many to many	24,579	37.68	29,500	45.22	24,945	38.24
9) Non-tracted area	18,344	28.12	8,298	12.72	—	—

Source: Tabulations from CensusCD Neighborhood Change Database.

Merging Other Data Sources With the NCDB

Merging other data with the NCDB can create an even more valuable and customized source of information. The ability to combine other geographic databases allows users to supplement the list of NCDB variables with those of their own. Nongeographic databases can also be appended to the NCDB. For example, users with survey data that contain respondents' home addresses could use the NCDB as a source of information on respondents' neighborhood characteristics and the opportunities or constraints they face at the local level.

Data cannot be merged with the NCDB using the software available on the NCDB CD-ROM but must be accomplished using some other data software—such as a database package (MS Access, dBase, FoxPro), data analysis software (SAS, SPSS, Stata), or mapping software (ArcView, ArcInfo, MapInfo). The actual merge procedure depends on the software being used. To merge the NCDB data with other sources, one must first export the appropriate NCDB

variables to an external file. The NCDB CD-ROM uses ASCII, dBase IV, ArcView Shape, and MapInfo Mid/Mif as export formats. These formats can be read by a wide variety of database, data analysis, and mapping software.

The external NCDB file must then be "merged," "linked," or "joined" to one or more other data files with the chosen software. This is accomplished by using a common identifier that exists in both files. Most likely, one will merge data to the NCDB by census tract identifier. When merging by tracts, remember that tract identifiers are unique only within U.S. counties. Therefore, when merging data from more than one county, use one of the "GEO" tract identifier variables, which include the state, county, and tract codes and is thus a unique tract identifier. The other data file will need to have an identically constructed variable to allow a successful merge with the NCDB data.

For some types of software, users may need to sort the observations in the data file by the geographic identifier before accomplishing the merge. It is also important to remember that the geographic identifiers in the NCDB are stored as character variables. If the corresponding identifier in the other data file is a numeric variable, users will most likely not be able to merge the two files successfully. Either create a character variable identifier in the non-NCDB data file or a numeric identifier in the NCDB file.

Finally, it is also possible to merge data from the NCDB at geographic levels other than tracts, such as state, county, or metropolitan area. In these cases, it is necessary to aggregate the NCDB data first to the appropriate level before attempting the merge. This will provide an NCDB file that is summarized with one observation for each state, county, or metropolitan area, depending on the geographic level at which the merge will be done. Most of the software described above for merging should also be able to summarize the data in this way.

Data Suppression

In accordance with federal law, information about individuals gathered in the decennial census must remain confidential. At first, this might not seem to be a problem for the NCDB, since data are aggregated at the tract level and no information is supplied about specific individuals. With the large number of complex cross-tabulations and the relatively small size of census tracts, however, it might be theoretically possible to derive information about certain individuals from census tabulations.

To illustrate, consider a tract with 4,000 people, 100 of whom are American Indians. If 50 of these American Indians were men, and 7 were over 65 years old, tables cross-tabulating race by age would provide information about these 7 identifiable individuals. For example, a table tabulating race by age by income that showed seven American Indian senior citizens living below the poverty level in the tract in question would reveal confidential income data about the seven individuals in question.

While such breaches of confidentiality may be unlikely, the Census Bureau must take steps to prevent them. Prior to 1990, the Bureau "suppressed" certain census data based on set criteria.[3] If, for example, the number of individuals in a particular tabulation cell fell below a set level, these data would not be reported. Therefore, some tracts in 1970 and 1980 have missing information due to suppression.

The Census Bureau places "flags" in its data to alert users to data suppression. The NCDB contains a set of similar, but not identical, flags to accomplish the same purpose. These flags are defined in relation to the original census source tabulations. So a user must find the appropriate flag by comparing the table source for the NCDB variables and then looking up the corresponding flag variable.

The NCDB suppression flags for particular 1970 and 1980 variables can be found in the data dictionary (appendix E). Suppression flags are character variables, coded as either blank (“ ”), to indicate no data suppression, or one (“1”), to indicate data suppression in one or more tabulation cells that make up that variable.

Undercount and Inaccurate Responses

Since its inception in 1790, controversy has surrounded the decennial census's alleged undercount of individuals (Anderson 1988). This is a significant issue because data from the census are so widely used in social science research and are the basis of important political decisions, including the drawing of congressional districts and the allocation of government funding. Today, critics of the census also point to the disproportionate undercount of racial and ethnic minorities, particularly young black men living in urban areas (Skerry 1992, West and Fein 1990).

No one, not even the Census Bureau, denies that the census misses many people. Also, to a lesser extent, there is some enumeration of fictitious or deceased individuals and double

counting. The undercount problem exists for many reasons. For instance, the Census Bureau may miss some housing units when sending out forms or some people who have received forms may not complete and return them. The former case is prevalent among individuals with no stable address (such as the homeless), while the latter is particularly common among illegal immigrants, many of whom wish to remain hidden from the government. While the Census Bureau makes several attempts to locate nonresponding households, some are inevitably missed.

Based on follow-up studies and comparisons with other data sources, the Census Bureau and others have estimated the overall undercount in recent censuses. It is generally thought that the number of missed individuals has fallen from around 5 percent in 1950 to under just under 2 percent in 1990 (Skerry 1992). Census 2000 is claimed to be one of the most accurate ever, although a definitive measure of the undercount has not yet been issued.[4]

Of particular concern is the so-called “differential undercount,” which refers to the fact that certain types of individuals and households are more likely to be missed by the census than others. According to one study, the undercount for black persons remained at 5.7 percent in 1990—an improvement from the 8.4 percent mark in 1940, but an increase from 4.5 percent in 1980 (Robinson, et. al. 1991). Men and the young are more likely to be missed than women and the old, and one study estimated that for black males between 20 and 29, the undercount was 10.1 percent in 1990 (Skerry 1992). The number of illegal immigrants, most of whom are of Hispanic origin, is believed to be around 3 million, and the Census Bureau estimates that 30 percent of this population was missed in 1990.

False or missing information from the census is not only an issue in counting individuals. One recent study contended that the census significantly overstates the number of female-headed families, particularly among blacks, because many male cohabitants are hidden to protect the household's welfare benefits (Hainer, et. al. 1988). Similar misreporting might be expected in other areas related to eligibility for public assistance, such as employment, disability, and earnings.

Proposals to "adjust" both the 1990 and 2000 census to correct for undercounting generated contentious debate. In July 1991, amid cries of political partisanship, Commerce Secretary Robert Mosbacher declined to make any adjustments to the official census numbers to correct for a possible undercount. Although admitting to the undercount and the

disproportionate rates among minorities, Mosbacher claimed the original count remained the fairest and best source of information on the nation's population (Elving 1991). More recently, the Executive Steering Committee for Accuracy and Coverage Evaluation Policy (ESCAP) recommended against the use of adjusted Census 2000 numbers for both redistricting and nonredistricting purposes (ESCAP 2001).

Users of the NCDB should recognize the problems associated with the acknowledged undercount of individuals. This phenomenon is particularly important for users who focus on urban poverty and the neighborhoods where undercounting is thought to be most common. Nevertheless, for all its shortcomings the census remains the only full count of persons in the United States, and for numerous topics it remains our best available data source.

The 1990 Homeless Count

In 1990, for the first time, the Census Bureau attempted to count the nation's homeless population. On Shelter and Street Night ("S-Night") in March 1990, enumerators set out to count the number of individuals spending the night in homeless shelters and those visible in predetermined street locations. This one-night project estimated the homeless to number around 230,000, including 179,000 in shelters and 50,000 on the street. Homelessness experts criticized the numbers as gross understatements of the population. Some advocates claim the actual number to be in the millions, and the Census Bureau has publicly admitted that the S-Night count missed many homeless individuals.

The variable HOMLES9N in the 1990 NCDB reports the total number of homeless individuals in shelters and visible in street locations. When comparing this number to other possible estimates of the homeless population, it should be remembered that the census S-Night count is a “point-in-time estimate,” meaning it counted people who were in shelters or on the street on one particular night. Other estimates may be based on whether people have been homeless at any point during an extended period, such as a month or year. These two types of estimates are not comparable. Furthermore, other estimates of homelessness may include families living in overcrowded housing or lacking permanent shelter of their own.[5]

Race Bridging

A major change in Census 2000 from previous censuses was the addition of multiracial categories in the collection and tabulation of the data. Respondents in Census 2000 were allowed to select one or more of six racial groups: White, Black/African American, Native American/Alaskan Native, Asian, Native Hawaiian/Other Pacific Islander, and "some other race." In previous censuses, respondents could choose only one racial group. Only about 2.4 percent of respondents nationwide selected more than one racial group in Census 2000, although this proportion was much higher in certain census tracts.

In tabulating population by race from the Census 2000 short form, the Census Bureau provided counts for all 63 combinations of the six racial groups that a respondent could have selected. To facilitate comparisons with previous censuses, the "race bridging" variables in the NCDB take all of the multiracial categories for Census 2000 and reapportion them into single racial groups. This allows one to compare racial change for tracts between the 1970, 1980, 1990, and 2000 censuses. There are many possible methods for creating bridging variables. The method we selected was developed by Jeffrey Passel of the Urban Institute's Population Studies Center.[6] It assigns multiracial groups to single races according to the rules below, in descending order of priority:

1) Black + any other race, assign to Black, otherwise

2) Asian + any other race, assign to Asian, otherwise

3) Native Hawaiian/Other Pacific Islander (NH/OPI) + any other race, assign to NH/OPI, otherwise

4) White + any other race, assign to White, otherwise

5) American Indian/Alaskan Native (AI/AN) + any other race, assign to AI/AN, otherwise

6) Assign to "Some other race"

For the sixth group, "Some other race," only people selecting this alone are assigned to that bridging category.

In addition to the race question, a separate "ethnicity" question asks each respondent whether they consider themselves to be Hispanic or Latino. This is similar to the way this

question was asked in earlier years, so no special method is needed for comparing these data across the censuses. (See the next section for more information.)

The NCDB 2000 variables with names starting SHR are the race bridging variables. There are matching sets of variables for most of these in previous years. The exceptions are the Asian and Native Hawaiian/Other Pacific Islander (NH/OPI) variables, which are not available separately in the NCDB for 1990, but only as a combined group—Asian/Pacific Islanders.

For 2000, there is also a series of NCDB variables that allow more analysis of the multiracial data. For each 2000 SHR variable, there is a corresponding MIN variable, which is the number of people who chose that race alone, and a MAX variable, which is the number of people who selected that race alone or in combination with another race. The bridging estimate must fall between these two numbers:

$$MIN \leq SHR \leq MAX$$

The MIN and MAX values therefore represent a range of possible bridging estimates for a population, so these variables can be used to examine the sensitivity of population by race changes to the particular bridging methodology used in the NCDB. If the range is very large, for instance, one might want to report some alternate methods of comparing racial groups in earlier years.

There is also a series of multiracial count variables with names starting MR1, MR2, MR3, and MRA, which are persons selecting one race only, two races only, three or more races, and multiple (i.e., two or more) races, respectively.

Because the determination of bridging status depends on the detailed multiracial counts from the short form (SF1) tabulations, the short form data had to be incorporated and adjusted to be compatible with the long form data. This is discussed further in the section “Incorporating 2000 Short Form Counts” in this chapter.

Figure 4-3: Percent Population by Race for 1990 and 2000

	1990	2000		
		Single/Multi Race	Single Race Only	Bridged
Total Persons (000s)	248,710	281,422	274,151	281,422
Percent Persons				
Total	100.0	100.0	100.0	100.0
White	80.3	75.1	77.1	76.3
Black	12.0	12.2	12.5	12.9
Am. Indian	0.8	0.9	0.9	0.9
Asian/PI	2.9	3.8	3.8	4.4
Other	3.9	5.5	5.6	5.5
Multiracial	—	2.6	—	—

Sources: CensusCD Neighborhood Change Database.

Figure 4-3 summarizes the population by race for 1990 and for different types of tabulations for 2000. The first two 2000 columns are based on standard census tabulations. The first column for 2000 ("Single/Multi Race") reports the percentages based on people who selected only a single race (in the NCDB, the MIN variables) and those who selected multiple races. The second column ("Single Race Only") eliminates the multiracial group and shows percentages based only on those picking a single race. The last column ("Bridged") shows the percentages calculated from the NCDB bridged data (the SHR variables).

For other variables broken down by racial groups (such as homeownership by race), we have not developed a "bridged" version in 2000 but have only reported the breakdowns by each race alone, the multiracial group, and Hispanic/Latino, following the Census Bureau tabulation of these items in the SF3. We decided not to develop bridged versions of these fields for two reasons. First, it would have greatly increased the number of variables in the NCDB file and have potentially created problems to fit the entire database on a single CD. Second, we felt the increase in the number of fields would have added more confusion for users trying to navigate the plethora of variables.

Therefore, apart from the bridged population counts, counts of persons and households by race are not directly comparable between 2000 and earlier census years. Such comparisons

can still be made, but users should examine the size of the multiracial population for the characteristic in question to make sure that this group is not distorting any comparisons to earlier years.

If users wish, they may construct their own bridged versions of race tabulations in 2000 by applying the information available in the NCDB. For example, to create a bridged version of the number of White homeowners in 2000 (OWNOCCW0), the following formula can be used:

Bridged OWNOCCW0 =
OWNOCCW0 +
((SHRWHT0N – MINWHT0N) / MRAPOP0N) × OWNOCCM0

This formula reapportions the multiracial homeowners (OWNOCCM0) according to the ratio of the multiracial population assigned to the White racial group (SHRWHT0N – MINWHT0N) over the total multiracial population (MRAPOP0N). This same formula can be used to create bridged variables for the other racial groups by substituting OWNOCCW0, SHRWHT0N, and MINWHT0N with the appropriate variables for those races.

Determining Hispanic/Latino Origin

In addition to race, Hispanic/Latino origin, or "ethnicity," is one of the major characteristics by which data in the census are tabulated. Tabulations by Hispanic/Latino origin are separate from race, since both racial affiliation and ethnicity have been identified for all persons enumerated in the census since 1980. In other words, Hispanic/Latino persons may declare themselves to be White, Black, Asian, Native Hawaiian/Pacific Islander, American Indian, or some other race, depending on the race options available for the particular census.

The terminology used to refer to persons of Hispanic/Latino ethnicity have changed from census to census. In the NCDB data and documentation, we use the term "Hispanic/Latino" to refer to all of these different ethnicity classifications across the four censuses. For the 1970 5-percent sample, a new census question identified persons of "Spanish origin or descent," which included Mexican, Puerto Rican, Cuban, Central or South American, or other Spanish. These persons were counted as "Spanish Americans" by the Census Bureau. For the 1970 15-percent sample, however, the question on Spanish origin was not asked and so the Census Bureau identified Spanish Americans from this sample according to a set of post-enumeration procedures depending on the region of nation. These were:

1) in New York, New Jersey, and Pennsylvania, persons born in Puerto Rico and their children (according to the Census documentation, persons of "Puerto Rican stock") were enumerated as Spanish American;

2) in Arizona, California, Colorado, New Mexico, and Texas, persons reporting Spanish as their mother tongue or as the language spoken by the wife or head of their family were counted as Spanish Americans, as were individuals whose surnames matched a list of 8,000 Spanish-American surnames; and

3) in the remaining states, individuals reporting Spanish as their mother tongue or the language spoken by the wife or head of their family were counted as Spanish Americans.[7]

In 1980, the question on "Spanish/Hispanic origin or descent," which was specifically defined as Mexican, Mexican-American, Chicano, Puerto Rican, Cuban, and other Spanish/Hispanic, was asked of all persons enumerated in the census for the first time. For the 1990 decennial census, the data on "Spanish/Hispanic origin" were derived from answers to questionnaire item 7, which was again asked of all persons. Persons of Hispanic origin were those who classified themselves in one of the specific Hispanic origin categories listed on the questionnaire—Mexican, Puerto Rican, and Cuban—as well as those who indicated that they were of other Spanish/Hispanic origin. In 2000, the ethnicity question was reworded yet again to identify persons who considered themselves to be "Spanish/Hispanic/Latino," more specifically defined on the questionnaire as Mexican, Mexican-American, Chicano, Puerto Rican, Cuban, and other Spanish/Hispanic/Latino.

The availability of data tabulated by Hispanic/Latino origin also varied greatly from census to census. In both 1970 and 1980, tabulations of data based on ethnicity were quite limited. In 1990 and 2000, census tabulations reporting data by race were usually followed by tabulations of the same data for all persons of Hispanic/Latino origin. In some cases, data was also further cross-tabulated by both race and ethnicity, that is, non-Hispanic/Latino White, non-Hispanic/Latino Black/African American, etc. The 2000 census SF3 tabulations included non-Hispanic/Latino White as a separate category for all race tabulations, whereas in the 1990 STF3 cross-tabulations by race and ethnicity were much less common.

Incorporating 2000 Short Form Counts

Prior to 2000, all tabulations available for short form data were generally replicated in the long form tabulations for that year and therefore all NCDB variables could be derived strictly from the long form tabulations. In 2000, however, several series of tabulations that were prepared for the short form (SF1) data were *not* available in the long form (SF3) tabulations. These included the detailed tabulations of the multiracial population, the detailed tabulations of Spanish origin (Mexican, Puerto Rican, Cuban, etc.), and the detailed group quarters counts. Since these tabulations were all required for creating certain NCDB variables, it was necessary to incorporate both short form and long form counts into the 2000 NCDB data.[8]

The problem is that short form and long form counts for particular geographic areas or subpopulations do not necessarily agree with each other, since the former are based on an enumeration of the entire population and the latter only on weighted sums of a 1-in-6 sample of households. As further explained by the Census Bureau:

> The differences between the long form estimates in SF 3 and values in SF 1 or SF 2 are particularly noticeable for the smallest places, tracts, and block groups. The long form estimates of total population and total housing units in SF 3 will, however, match the SF 1 and SF 2 counts for larger geographic areas such as counties and states, and will be essentially the same for medium and large cities.[9]

Our method for incorporating the short form counts into the long form data had three steps. First, we created the appropriate short form variables directly from the SF1 tabulations for every 2000 census tract. Second, we used control totals from the long form to "ratio adjust" the short-form-derived variables so that they would agree with the long form tract totals. Finally, we rounded the ratio-adjusted short form tract counts to whole numbers and subtracted any residual so that the new values added up to the correct control total.

For instance, one set of variables derived from the short form was the bridged race variables described earlier: SHRWHT0N, SHRBLK0N, etc. These variables must be derived from the detailed multiracial tabulations available only in the short form data. In fact, these variables can be expressed as the sum of two counts, the multiracial population assigned to the race category and the single race population that chose that same race. It is actually the former that we want to adjust since the latter has an equivalent long form count available. Therefore, we calculated the multiracial population assigned to each race category as:

MRAWHT0N = SHRWHT0N – MINWHT0N
MRABLK0N = SHRBLK0N – MINBLK0N
etc.

The variables MRAWHT0N, MRABLK0N, etc. do not exist in the NCDB file but were calculated for the purpose of this exercise. The variables SHRWHT0N, SHRBLK0N, etc. were derived from the short form tabulations; MINWHT0N, MINBLK0N, etc. were derived from the long form tabulations. Once these variables were created, we compared them to the long form total multiracial population (MRAPOP0N) according to the following formula:

MRAPOP0N = MRAWHT0N + MRABLK0N + MRAAMI0N + MRAASN0N + MRAHIP0N + MRAOTH0N

The population total, SHR0D, is the long form control total. Each of the terms on the right-hand side of the equation above were multiplied by the ratio:

(MRAWHT0N + MRABLK0N + MRAAMI0N + MRAASN0N + MRAHIP0N + MRAOTH0N) / MRAPOP0N

This ratio adjusted the short form variables so that they added up to the control total. Since NCDB counts must be whole numbers, the ratio-adjusted values were rounded up or down to the nearest integer. Since this rounding may cause the short form variables to no longer add up exactly to the control total, the largest value among the right-hand side variables was adjusted up or down to reconcile the two sides of the equation.

Ratio adjustments of short form counts were carried out for the non-Hispanic/Latino population by race (SHRNHW0N, SHRNHB0N, etc.) in a similar manner. Other fields were adjusted according to the following control totals:

MRAPOP0N = MR2POP0N + MR3POP0N

SHRHSP0N = MEXIC0 + PRICAN0 + CUBAN0 + DOMIN0 + COSRIC0 + GUATEM0 + HONDUR0 + NICARAG0 + PANAMA0 + SALVAD0 + OTHCAX0 + ARGNTN0 + BOLIVA0 + CHILE0 + COLOMB0 + ECUAD0 + PARAGY0 + PERU0 + URGUAY0 + VENZUL0 + OTHSAX0 + OTHHISP0

INSTP0N = CORR0N + AGED0N + MENTL0N + JUV0N + OINST0N

NOINP0N = DORM0N + MILTQ0N + ONINS0N

Comparing Monetary Values Across Censuses

The Census Bureau reports dollar values for income and housing costs in nominal or current dollars—the value in the year dollars were spent or earned. If an indicator like median income is compared across time using nominal dollars, the percentage change will reflect two factors: 1) the real change in purchasing power and 2) inflation, which is the overall general upward price movement of goods and services.

To measure the real change, the nominal dollar values must be converted or "deflated" to constant or inflation-adjusted dollar values. One of the most commonly used deflators is the Consumer Price Index (CPI). Produced by the Bureau of Labor Statistics (BLS), the CPI is a measure of the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services. The CPI and its components are used to adjust other economic series for price changes and to translate these series into inflation-adjusted dollars.

For the CPI, prices were set at a "base" of 100 between 1982 and 1984. Since the CIP is an index, it is the ratio of values between two years that is used to convert to constant dollars. To adjust for inflation, first select the "base" year (the year to which the dollars will be converted) and divide the index for the year in question into that base year. For example, in 1989 the median family income was $30,056 in unadjusted dollars. To adjust this amount to 1999 dollars, the 1999 annual average CPI value is divided by the average annual index value for 1989:

1.666 (1999 index value) / 1.240 (1989 index value) = 1.344

The answer, 1.344, becomes the multiplier to turn the 1989 median family income into its equivalent 1999 dollars: 1.344 × $30,056 = $40,382. For more information about the CPI and other BLS deflators, see http://www.bls.gov/bls/inflation.htm.

Notes

[1] This differs from the method used to remap the tracts in the original UDB. At that time, GIS technology was not readily available to allow the use of census block weightings. Instead, the remapping was done by using "tract correspondence files" provided by the Census Bureau. These files simply list the correspondences between tracts for two successive census years, but do not provide any geographical information or population weightings. Therefore, if a tract split into three pieces, for the UDB the

population would have been divided equally among all three parts. The availability of GIS technology and boundary files should greatly improve the quality of the tract remapping for the NCDB over the original UDB.

[2] In the UDB, the 1970 and 1990 data were remapped to 1980 tract boundaries. This was done to be consistent with the first version of the UDB developed by Sawhill and Ricketts, which contained only 1980 data.

[3] Starting with the 1990 census, the Census Bureau used "data swapping" as the preferred method for protecting the confidentiality of individual responses. Data swapping involves either editing the census source data or exchanging records for a sampling of cases when tabulating the data. The swapped records are matched on a set of key criteria and should not affect the accuracy of the data at higher levels of aggregation. For more information, see U.S. Bureau of the Census (1993, appendix C) and the confidentiality information on the American FactFinder web site (http://factfinder.census.gov/home/en/confidentiality.html#dataswapping).

[4] According to the report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy (ESCAP), the preliminary estimates of Census 2000 coverage range from an overcount of 0.65 percent to an undercount of 1.15 percent (ESCAP 2001, p. iv).

[5] For more on S-Night and follow-up studies evaluating its accuracy, contact the Center for Survey Methods Research at the Census Bureau. For additional information on estimates of the homeless population, see Burt et al. (1999).

[6] The Office of Management and Budget (OMB) provides guidance on using multiracial data for the purposes of civil rights monitoring and enforcement. See OMB (2000) for more information.

[7] In the 1970 source tabulations, Population Count Table 24 reports the results of all four procedures for allocating persons to the Spanish-American population.

[8] Counts derived from the Census 2000 short form are identified in the NCDB data dictionary with a source specification starting with "SF1:", for example, "SF1: Table P3:3,12,15,33."

[9] American FactFinder web site, http://factfinder.census.gov/servlet/DatasetMainPageServlet?_program=DEC&_lang=en&_ts=, "Comparing SF 3 Estimates with Corresponding Values in SF 1 and SF 2," accessed October 26, 2003.

References

Anderson, Margot. 1988. “Planning the Future in the Context of the Past.” *Society* 25 (March/April): 39–47.

Burt, Martha R., Laudan Y. Aron, Toby Douglas, Jesse Valente, Edgar Lee, and Britta Iwen. 1999. *Homelessness: Programs and the People They Serve: Findings of the National Survey of Homeless Assistance Providers and Clients.* Washington, D.C.: The Urban Institute.

Clark, Rebecca. 1992. *Neighborhood Effects On Dropping Out of School Among Teenage Boys.* Washington, D.C.: The Urban Institute. Discussion Paper #PSC-DSC-UI-13.

Crane, Jonathan. 1991. “The Epidemic Theory of Ghetto and Neighborhood Effects on Dropping Out and Teenage Childbearing.” *American Journal of Sociology.* 96: 1226–1259.

Ellen, Ingrid, and Margery Turner. 1997. “Location, Location, Location: How Does Neighborhood Environment Affect the Well-Being of Families and Children?” *Housing Policy Debate.* 8(4): 833–866.

Elving, Ronald D. 1991. “Refusal To Adjust Undercount Spurs Protest, Renews Suit.” *Congressional Quarterly*. 49 (July 20): 2006–2009.

Executive Steering Committee for Accuracy and Coverage Evaluation Policy (ESCAP). 2001. *Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy on Adjustment for Non-Redistricting Uses.* Washington, D.C.: U.S. Census Bureau. http://www.census.gov/dmd/www/EscapRep2.html. (Accessed April 30, 2002.)

Federal Information Processing Standards (FIPS). 1995. *Announcing the Standard for Metropolitan Areas, including MAs, CMSAs, PMSAs, and NECMAs* Washington, D.C.: National Institute of Standards and Technology. (FIPS Publication 8–6). http://www.itl.nist.gov/fipspubs/fip8–6–0.htm. (Accessed May 7, 2002.)

Galster, George, Ronald Mincy, and Mitch Tobin. 1993. “The Disparate Neighborhood Impacts of Economic Restructuring.” Paper prepared for presentation at the Allied Social Science Association/National Economics Association Meetings, Boston, Mass. Hainer, Peter, Catherine Hines, Elizabeth Martin, and Gary Shapiro. 1988. “Research on Improving Coverage in Household Surveys.” *Fourth Annual Research Conference Proceedings.* Washington, D.C.: U.S. Census Bureau.

Massey, Douglas, and Nancy Denton. 1993. *American Apartheid: Segregation and the Making of the Under Class.* Cambridge, Mass: Harvard University Press.

Mincy, Ronald. 1988. “Is There a White Underclass?” Washington, D.C.: The Urban Institute. Mimeo.

———. 1995. “Ghetto Poverty: Black Problem or Harbinger of Things to Come?” In *African American Economic Thought: Volume 2, Methodology and Policy*, edited by Thomas D. Boston. New York: Routledge.

Mincy, Ronald, and Susan Wiener. 1993. *The Under Class in the 1980s: Changing Concept, Constant Reality.* Washington, D.C.: The Urban Institute. Research Paper.

Office of Management and Budget. 1990. “Revised Standards for Defining Metropolitan Areas in the 1990s.” *The Federal Register.* 55(62): 12154–12160.

———. 2000. *Guidance on Aggregation and Allocation of Data on Race for Use in Civil Rights Monitoring and Enforcement* (OMB Bulletin No. 00–02). http://www.whitehouse.gov/omb/bulletins/b00–02.html. (Accessed April 30, 2002.)

Ricketts, Erol, and Ronald Mincy. 1989. “Growth of the Underclass.” *Journal of Human Resources*. 25(1): 137–145.

Ricketts, Erol, and Isabel Sawhill. 1988. “Defining and Measuring the Underclass.” *Journal of Policy Analysis and Management*. 7(2): 316–325.

Robinson, J. Gregory, Bashir Ahmed, Prithwis Das Cupta, and Karen Woodrow. 1991. “Estimating Coverage of the 1990 Census: Demographic Analysis.” Paper presented at meeting of the American Statistical Association, Aug. 20 (cited in Skerry, 1992).

Skerry, Peter. 1992. “The Census Wars.” *The Public Interest*. 106 (Winter): 17–31.

Tobin, Mitchell. 1993. “Poverty and the Under Class in States and Metropolitan Areas: 1990.” Washington, D.C.: The Urban Institute. Mimeo.

———. 1993b. *Users’ Guide for the Urban Institute’s Under Class Data Base (UDB).* Washington, D.C.: The Urban Institute.

U.S. Bureau of the Census. 1970. *1970 Census Users' Guide, Parts I and II*. Washington, D.C.: U.S. Department of Commerce.

———. 1982. *Census of Population and Housing, 1980: Summary Tape File 3, Technical Documentation.* Washington, D.C.: U.S. Department of Commerce.

———. 1993. *Census of Population and Housing, 1990: Summary Tape File 3 Technical Documentation.* Washington, D.C.: U.S. Department of Commerce.

U.S. Census Bureau. 1994. *Geographic Areas Reference Manual.* Washington, D.C.: U.S. Department of Commerce. http://www.census.gov/geo/www/garm.html (Accessed February 21, 2001.)

———. 2001. *Census 2000 Summary File 1 Technical Documentation.* Washington, D.C.: U.S. Department of Commerce.

West, Kirsten K., and David J. Fein. 1990. "Census Undercount: An Historical and Contemporary Sociological Issue." *Sociological Inquiry.* 60(2): 127–141.

Wilson, William Julius. 1987. *The Truly Disadvantaged*. Chicago: The University of Chicago Press.

Zimmerman, Wendy, and Mitchell S. Tobin. 1995. *Immigration and Concentrated Poverty.* Washington, D.C.: The Urban Institute.