Data on Health and Well-being of American Indians, Alaska Natives, And Other Native Americans

Data Catalog

Contract No. 233-02-0087

Prepared for:
U.S. Department of Health and Human Services
Office of the Assistant Secretary for Planning and Evaluation
Office of Human Services Policy
200 Independence Avenue, S.W.
Washington, D.C. 20201

Prepared by:
Westat
1650 Research Boulevard
Rockville, MD 20850

December 2006

This report was produced under the direction of Alana Landey and Peggy Halpern, Task Order Officers, Office of the Assistant Secretary for Planning and Evaluation, Office of Human Services Policy, Jerry Regier, Principal Deputy/Assistant Secretary for Planning and Evaluation.
This report is available online at:

http://aspe.hhs.gov/hsp/06/Catalog-AI-AN-NA

Acknowledgments

This catalog was prepared by staff from Westat, the American Indian Health Research Program of Black Hills State University, and L&M Policy Research, LLC. Peggy Halpern and Alana Landey of the Office of the Assistant Secretary of Planning and Evaluation (ASPE), United States Department of Health and Human Services (DHHS), served as Task Order Officers for this effort and provided input and guidance to the project. Canta Pian, director of the Division of Economic Support for Families in ASPE's Office of Human Services Policy, provided review and comment on this report.

Three members of the DHHS Data Council's Racial and Ethnic Data Working Group—Audrey Burwell, Dale Hitchcock, and Edna Paisano—served on a project workgroup and provided guidance to the project as well as review of this catalog. Several members of an American Indian, Alaska Native, and other Native American workgroup also provided input and review, including David Wong and Ralph Bryan representing the CDC Office of Minority Health and Health Disparities and the IHS Division of Epidemiology and Disease Prevention, David Simmons representing the National Indian Child Welfare Association, and Lisa Oshiro representing the Council for Native Hawaiian Advancement.

We also benefited from input from several external consultants to the project, including Gordon Belcourt, Executive Director of the Montana-Wyoming Tribal Leaders Council; Carole Anne Heart, Executive Director of the Aberdeen Area Tribal Chairmen's Health Board; Jeff Henderson, President of the Black Hills Center for American Indian Health; Lilia Kapunai, Vice President of Council for Native Hawaiian Advancement; and Frank Ryan, independent consultant. Numerous knowledgeable individuals at DHHS and other federal and non-federal organizations gave generously of their time to ensure that we had complete and up-to-date information on existing surveys and databases. We appreciate the careful review of the data catalog and data source profiles conducted by all these individuals.

Table of Contents

Section **Page**

1. Background 1-1

2. Catalog Description 2-1
   2.1 Catalog Organization 2-3
   2.2 Methodology 2-5
   2.3 Using These Data Sources 2-7

3. Data Sources by Topical Area 3-1

4. Data Sources by Subpopulation Coverage 4-1

5. Listing of Data Sources 5-1
   A Picture of Subsidized Households (1998) 5-2
   Adoption and Foster Care Analysis and Reporting System (AFCARS) 5-4
   American Community Survey (ACS) 5-8
   American Housing Survey (2003) 5-12
   American Housing Survey: Metropolitan Surveys 5-15
   Annual Survey of Jails (ASJ) 5-19
   Behavioral Risk Factor Surveillance System (BRFSS) 5-22
   CAHPS Health Plan Survey Response Data 5-25
   California Health Interview Survey (CHIS) 5-29
   Census 2000 5-33
   Census 2000 - The American Indian and Alaska Native Summary File 5-38
   Census of Agriculture (2002) 5-41
   Census of Jails 5-44
   Census of Tribal Justice Agencies in Indian Country (2002) 5-47
   Consumer Expenditure Surveys (CE) Interview and Diary Surveys 5-49
   Current Population Survey (CPS) 5-53
   Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) 5-58
   Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K) ... 5-63
   Food Stamp Quality Control Database (FSPQC) 5-67
   Hawaii Health Survey (HHS) 5-70
   Head Start Program Information Report 5-73
   Health and Retirement Study (HRS) 5-75
   Health Behavior in School-aged Children (HBSC) 5-78
   Health Information National Trends Survey (HINTS) 5-81
   Integrated Postsecondary Education Data System (IPEDS) 5-83
   Medicaid Analytic Extract (MAX) 5-85
   Medical Expenditure Panel Survey (MEPS) 5-88

Table of Contents (continued)

Section **Page**

Medicare Denominator File 5-92
Medicare Utilization – Standard Analytic Files (SAFs) 5-94
National Aging Program Information Systems (NAPIS) State Program Reports 5-96
National Ambulatory Medical Care Survey (NAMCS) 5-98
National Assessment of Adult Literacy (NAAL) 5-101
National Child Abuse and Neglect Data System (NCANDS) 5-104
National Crime Victimization Survey (NCVS) 5-108
National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) . 5-111
National Health Interview Survey (NHIS) 5-114
National Hospital Ambulatory Medical Care Survey (NHAMCS) 5-117
National Household Education Surveys Program (NHES) 5-120
National Household Travel Survey (NHTS) 5-125
National Indian Education Study (NIES) 5-130
National Longitudinal Mortality Study (NLMS) 5-133
National Mortality Followback Survey (NMFS) 5-135
National Survey of America's Families (NSAF) 5-137
National Survey of Child and Adolescent Well-being (NSCAW) 5-142
National Survey of Family Growth (NSFG) 5-146
National Survey of Veterans (NSV) 5-150
National Survey on Drug Use and Health (NSDUH) 5-153
National Vital Statistics System: Linked Birth-Infant Death (NVSS-I) 5-156
National Vital Statistics System: Mortality (NVSS-M) 5-159
National Vital Statistics System: Natality (NVSS-N) 5-162
Panel Study of Income Dynamics (PSID) 5-164
Pediatric Nutrition Surveillance System (PedNSS) 5-167
Pregnancy Nutrition Surveillance System (PNSS) 5-169
Pregnancy Risk Assessment Monitoring System (PRAMS) 5-171
Resource and Patient Management System (RPMS) 5-174
Runaway and Homeless Youth Management Information System (RHYMIS) 5-177
Small Area Income and Poverty Estimates (SAIPE) 5-180
Surveillance, Epidemiology, and End Results (SEER) 5-183
Survey of Jails in Indian Country (SJIC) 5-186
Survey of Program Dynamics 5-188
Temporary Assistance for Needy Families (TANF) and Tribal TANF 5-191
Tobacco Use Supplement to the Current Population Survey (TUS-CPS) 5-194
Treatment Episode Data Set (TEDS) 5-199
Uniform Crime Reports (UCR) 5-202

Table of Contents (continued)

Section **Page**

United States Renal Data System (USRDS) 5-205
Washington State Population Survey (WSPS) 5-207
Youth Gangs in Indian Country 5-209
Youth Risk Behavior Surveillance System (YRBSS) 5-211

6. Other Data Sources 6-1

7. Other Reports 7-1

Appendices

Appendix A: Glossary A-1

Appendix B: Data Set Aggregation B-1

Appendix C: Review Protocol C-1

Appendix D: List of Sponsoring Agencies D-1

Figure and Tables

Figure 1.1 Key Data Catalog Policy Issues 2-2

Table 1. Data Sources by Policy Issue 3-1
Table 2. Data Sources by Subpopulation Coverage: American Indian/Alaska Native 4-1
Table 3. Data Sources by Subpopulation Coverage: Native Hawaiian/Other Pacific Islander 4-4

Data on Health and Well-being of American Indians, Alaska Natives and Other Native Americans: Data Catalog

1. Background

The Study of Data on Health and Well-being of American Indians, Alaska Natives, and other Native Americans (AI/AN/NAs) was funded by the Department of Health and Human Services' (DHHS) Office of the Assistant Secretary for Planning and Evaluation (ASPE) to address the need for systematic information about available data sources pertaining to the health and well-being of AI/AN/NA populations. This study examined numerous existing databases—federal surveys, research survey databases, state and community surveys, and administrative databases—that include information on AI/AN/NA population characteristics and measures of health and well-being. The study team documented the nature of these databases, including their strengths and limitations, and collated the information into this data catalog. In the course of a systematic review, the study also shed light on the limitations and gaps in available data on the health and well-being of AI/AN/NA populations. The second component of this project, a paper entitled *Report on Gaps in Data, Initiatives Underway, and Strategies for Improving AI/AN/NA Data for Policy and Research* describes these limitations and gaps and identifies possible strategies to improve the quality, usefulness, and population and geographic coverage of data on AI/AN/NA health and well-being.[1]

This study continues DHHS' focus on improving data collection concerning the health and well-being of racial and ethnic populations. The current study builds on previous activities, including the 1999 Joint Report of the DHHS Data Council's Working Group on Racial and Ethnic Data and the Data Workgroup of the DHHS Initiative to Eliminate Racial and Ethnic Disparities in Health entitled *Improving the Collection and Use of Racial and Ethnic Data in DHHS*. It also expands on the activities conducted during an earlier study for ASPE entitled *Assessment of Major Federal Data Sets for Analyses of Hispanic and Asian Pacific Islander Subgroups and Native Americans*. [2]

This data catalog focuses on data sources that provide information on the health and well-being of AI/AN/NA populations. The catalog is intended for use by a wide variety of users including AI/AN/NA communities; researchers from government, academic institutions, and foundations; and policy makers. The catalog provides overview information on possible data sources that could be used to describe the need for services, analyze trends in well-being and health, or illuminate disparities. Some of the data sources profiled in this catalog supply only published tables for the user. Others can be used by those with the necessary analytical skills and tools to do analysis on specific questions. The profiles of data sets presented in this catalog are not meant to provide instruction for use of the data in addressing specific research questions, but instead to serve as a source of general information that will help potential users determine if further investigation of a data source is warranted. The catalog also provides contact information and data source locations for conducting further in-depth reviews of targeted data sources.

The populations covered by this catalog are American Indians, Alaska Natives, and other Native Americans including Native Hawaiians and other Pacific Islanders. While each data source profiled here may collect information on the race of the respondents differently, it is helpful to keep in mind some generally accepted definitions of the key racial groups included in this catalog.

---

[1] This paper will be available online in the future.

[2] Reports from this effort are available at http://aspe.hhs.gov/datacncl/racerpt/ and http://aspe.hhs.gov/HSP/minority-db00/task2/index.htm.

American Indians (AI) and Alaska Natives (AN) are defined by the U.S. Census Bureau as "people having origins in any of the original peoples of North and South America (including Central America), and who maintain tribal affiliation or community attachment."[3] For purposes of measuring, monitoring, and developing approaches to reducing disparities in health and well-being and for research on a range of issues related to health and well-being, there may be interest in information on the combined AI/AN group for some purposes and in information on specific subgroups of the population for other purposes. For this reason, this data catalog includes information on the availability of data on the combined AI/AN category as well AI and AN as separate groups and AI/AN who are members of federally or state-recognized tribes. The catalog also notes the availability of information on those who self report that they are AI/AN alone or AI/AN in combination with other races. Finally, because there may be interest in identifying those American Indians who live on reservations, the reviewers examined each data source to determine if reservation of residence was available in that data source.

The Native Hawaiian (NH) and other Pacific Islander (PI) population includes those who are members of any of the native peoples of Hawaii and native peoples of Pacific Insular Areas that are dependent territories of the U.S. (Guam, American Samoa, and the Northern Mariana Islands) or Freely Associated States for which the U.S. provides defense, funding grants, and social services to its citizens (Micronesia, Palau, and the Marshall Islands). Within the NH/ PI population, there are several ethnically distinct categories. Polynesians, including Native Hawaiians, Samoans, Tongans, Tahitians, Tuvaluans, and Maori, are the largest group, accounting for 65 percent of all NH/PIs. Micronesians, including Guamanians, Marshallese, Palauans, residents of the Northern Mariana Islands and of the Federated States of Micronesia, are 13 percent of all NH/PIs. Melanesians, including Fijians, New Caledonians, Solomon Islanders, Vanuatuans, and Papua New Guineans, are 2 percent of this population.[4] In 1997, the Office of Management and Budget established a new racial category, Native Hawaiian and Pacific Islander (NH/PI), and required that federal agencies collect information on this new race category by 2003. The 2000 Census included the NH/PI race category and, as a result, provides an initial baseline for assessing socioeconomic status and some limited health measures of this group.[5]

---

[3] U.S. Census Bureau, "The American Indian and Alaska Native Population: 2000." Census 2000 Brief, February 2002, page 2.
[4] Panapassa, S.V. "The Health of U.S. Pacific Islander Populations: Emerging Directions." Presentation, May 2005.
[5] *Ibid.*

2. Catalog Description

As described above, this catalog is meant to provide overview information on a wide variety of data sources that can address health and well-being issues for AI/AN/NA populations. It is not meant to be an exhaustive listing of all data available on AI/AN/NA populations concerning health and well-being. Time and resource limitations prevented coverage of the entire universe of data sources that might be used to address these topics. To ensure that the catalog would provide broad coverage of the major topics of health and well-being, as a first step, the project staff, in consultation with ASPE, representatives of the DHHS Data Council's Racial and Ethnic Data Working Group (a workgroup for this project), and a small AI/AN/NA workgroup developed a detailed list of policy issues within the categories of health and well-being that should be covered in the catalog. The purpose of this policy list was to guide decision making about the content of the data sources that should be included. The project staff attempted to maximize coverage of the policy issues and avoid redundancy in the data sources being reviewed. This list of policy issues is presented in Figure 1.

Beyond policy parameters, the project team, in consultation with ASPE, members of the DHHS workgroup, and the AI/AN/NA workgroup, also established the following technical parameters for the data catalog:

- Data should be from a survey, program reporting system, or registry;
- Data should be quantitative in nature;
- Data source must allow user to identify people who are AI/AN/NA or focus on a specific geographic region with a large AI/AN/NA population;
- Data source must be available to researchers (as data or as requested analyses) or have extensive published tables and reports available;
- Contact and location information regarding the data source must be available;
- Documentation of data collection methods should be available;
- Data should not have been collected primarily to fill advocacy needs, lobbying purposes, or to support corporate interests;
- Total unweighted AI/AN/NA population in data source should be available, and the sample size must be adequate for analyses;
- Coverage of data source must be either national or focus on a specific AI/AN/NA subpopulation only (e.g., single tribe, Pacific Islanders), or consist of a smaller geographic area of clear relevance to the AI/AN/NA population; and
- Timeframe of the data source should be mid-1990s or later (unless a strong argument can be made to include older data).

The review of data sources conducted based on these parameters did not include an assessment of the quality of the data source. A careful examination of the actual data to make such an assessment was beyond the scope of this effort. Instead, reviewers sought to provide sufficient information about the data source to allow users to make an initial assessment of the potential usefulness of the data source for their purposes. We strongly recommend that potential users thoroughly examine the documentation and data from these sources to make their own assessments of the quality of the data.

Figure 1.1 Key Policy Issue Areas Guiding Inclusion of Data Sets

DEMOGRAPHIC AND ECONOMIC INDICATORS (e.g., age distribution, marital status, household composition)

HEALTH POLICY ISSUES

1. Measurement of health status (e.g., self-reported health, disability rates, mortality/morbidity rates, trends over time)
2. Disease-specific measurements (e.g., % with diabetes, TB, STDs, cancer)
3. Key health disparities of priority interest (e.g., prenatal care/birth outcomes, cancer mortality, substance abuse, alcohol use, mental health, suicide)
4. Factors contributing to measured health disparities (e.g., access to health care, utilization rates, insurance coverage, health care financing, socioeconomic factors, preventative measures (such as immunization rates))
5. Identification of evidence-based practices and programs that address causes of health disparities, result in positive health outcomes, and are generalizable/replicable
6. Role of traditional medicine in AI/AN/NA communities

WELL-BEING ISSUES

Economic Well-being

1. Income status (e.g., household income/poverty status, per capita income)
2. Unemployment rates
3. Economic assistance program participation rates (e.g., Temporary Assistance for Needy Families/Tribal Temporary Assistance for Needy Families, Food Stamps)
4. Economic opportunity (e.g., number of businesses/jobs, work history)
5. Measurement of economic/employment disparities between AI/AN/NA and general population
6. Factors contributing to economic disparities (e.g., lack of child care arrangement, transportation barriers)
7. Identification of evidence-based practices and programs that reduce economic disparities and are generalizable/replicable

Education Levels and Opportunities

1. Educational attainment (e.g., last grade completed, literacy/numeracy skills)
2. Educational opportunities (e.g., Head Start, special education programs, school financing)
3. Factors contributing to educational disparities (e.g., parents' education level, average education in city/county, education spending per capita, and other socioeconomic factors)
4. Identification of evidence-based practices and programs that produce positive educational outcomes and are generalizable/replicable

Family Well-being

1. Measures of well-being for families/households (e.g., families with low income levels, homeless families, teen pregnancy/birthrates, household size and composition)
2. Factors contributing to well-being disparities of families (e.g., socioeconomic factors, education levels of family adults, housing quality, public transportation availability)
3. Identification of evidence-based practices and programs that improve family well-being and are generalizable/replicable

Child Well-being

1. Measures of well-being for children (e.g., children in foster care, incarcerated children)
2. Factors contributing to well-being disparities of children (household composition, martial status of parents, foster care placement)
3. Identification of evidence-based practices and programs that improve child well-being and are generalizable/replicable

Elder Well-being

1. Measures of well-being for elders (e.g., elders with low income levels, homeless elders, elder abuse)
2. Factors contributing to well-being disparities of elders (e.g., socioeconomic factors, living arrangements, activities of daily living and instrumental activities of daily living (ADL/IADL), family members in proximity, services available/used (such as Meals on Wheels/elder transportation)
3. Identification of evidence-based practices and programs that improve elder well-being and are generalizable/replicable

Housing Issues

1. Housing quality (e.g., rooms per person, running water, electricity, heat, age of building)
2. Type of housing
3. Housing ownership
4. Rental unit quality and cost
5. Homelessness

Transportation Quality and Availability Issues

Justice System Issues

1. Rates of involvement with justice system (e.g., arrest, conviction, probation, parole rates)
2. Differences in resolution of arrest, by type of court system (e.g., federal, tribal, state, local)
3. Lifetime probability of being a victim of a violent crime
4. Lifetime probability of being a victim of a non-violent crime
5. Domestic violence rates
6. Child maltreatment rates
7. Factors contributing to disparities in involvement with justice system and outcomes (e.g., family stability/foster care placement, family members' history of legal system involvement, race/ethnicity, truancy history)
8. Identification of evidence-based practices or programs that reduce involvement with justice system or reduce recidivism and are generalizable/replicable

Military Service/Veterans' Issues

1. Military service rates (e.g., % served in military, % retired from military with benefits)
2. Eligibility and use of Veterans Administration health facilities
3. Eligibility and use of other Veterans Administration benefits (e.g. housing loans, educational benefits)

2.1 Catalog Organization

The data catalog is organized into six sections beyond this introduction. Each of these sections is described in the bullets below:

- **Section 3** provides a listing of the data sources included in the catalog sorted by key policy areas. The data sources listed in each section are particularly appropriate for answering questions in the topical area but may also contain data in other areas; therefore, data sources may be listed under more than one policy area.

  Under each policy topic, data source name and corresponding page number are presented. The page number can be used to find the detailed information on the data source in Section 5.
- **Section 4** provides a listing of data sources included in the catalog sorted by subpopulations identified and analyzable (e.g., all AI/AN/NA, American Indians and Alaska Natives, Pacific Islanders, Native Hawaiians, named tribes or communities). Page number is also included.
- **Section 5**, which is the largest section of this report, provides in depth profiles of all data sources presented alphabetically. A more detailed description of topics covered in these detailed profiles is presented below.
- **Section 6** includes a list of data sources identified but not reviewed for the catalog. This list is also alphabetical and is briefly annotated with reasons the data sources were not included in the review and selected information collected about the data set prior to its exclusion.
- **Section 7** includes reference information on a set of reports identified during this review that ASPE and the project team believe might be of particular interest to catalog users. These reports are generally of three types: reports concerning data that are not directly available to potential users but on which there are published detailed analytical reports, reports that are data compilations covering important issues using multiple data sources, and selected reports concerning the profiled data sources identified for review. This list of reports is not meant to be exhaustive but represents potentially useful information obtained in the course of assembling this catalog.
- **Appendices** to this catalog include a glossary of terms, a brief discussion of issues related to data aggregation, and the review protocol used to screen and profile the data sources.

As noted above, **Section 5** provides detailed profiles of each data source. Included in the detailed profiles are 25 possible elements. If they are not relevant to the data source, elements are not included in the profile; for example, if no information on AI/AN/NA subgroups is available, this field is not shown in the profile. The 25 possible elements for a data source profile include:

1. Name of Data Source
2. Sponsor—documents the agency or funding source for the data collection.
3. Description—provides a summary description of the purpose of the data collection and the general content of the data set.
4. Data Type(s) — indicates whether the data source is a survey, registry, or program reporting database.
5. Relevant Policy Issues—draws from the list of policy issues described above.

6. Unit of Analysis—describes the level(s) at which the data were collected (e.g., individuals, families, households, farms). This information can help users determine the type of analysis that can be conducted and the research questions that can be asked.
7. Identification of AI/AN/NA—provides detailed information on how data about race and ethnicity were collected for this data source. Where possible, question or form field text is included.
8. AI/AN/NA Population in Data Set—provides the unweighted count of American Indian, Alaska Native, or other Native American persons or families included in the data source. This information is provided in as much detail as is available (e.g., for American Indians only).
9. AI/AN/NA subpopulations—provides information on subpopulations for which analysis is possible (e.g., American Indian alone, tribal affiliation).
10. Geographic Scope—indicates the level of geographic analysis possible (e.g., national only, state, region).
11. Date or Frequency—provides information on the date of data collection and how frequently, if relevant, the data collections are repeated.
12. Aggregation— this section is included only where it is necessary or particularly relevant to consider combining data collections across years. For example, when the AI/AN/NA sample size is extremely small in one year, it may be necessary to combine multiple years of data to obtain a sample size appropriate for statistical analysis. Or, when the data are collected in segments across several years, it may be desirable to aggregate across the years to provide full population coverage.
13. Data Collection Methodology—provides some detail on how the data are collected (e.g., telephone, paper questionnaire.)
14. Participation—documents whether participation in the data collection effort was mandatory or optional, and may provide information on the use of incentives.
15. Response Rate—documents the response rate or coverage of the data collection. Unweighted response rates are provided where possible as these are the most directly interpretable. Where only weighted response rates are available, they are reported and, in most cases, briefly explained.
16. Sampling Methodology—provides detail on how individuals, household, or other foci were selected to participate in the data collection. This is usually only appropriate for surveys and, rarely, program reporting databases.
17. Oversample of AI/AN/NA Population—addresses the degree to which AI/AN/NA persons were purposely selected in excess of what would occur randomly.
18. Analysis—includes information important to conducting statistical analyses of the data; for example, information on the standard errors of any estimates, the effective sample size, design effects, and description of the power available to detect differences in the data.
19. Authorization—describes the authorizing legislation for the data collection, where appropriate.
20. Strengths—summarizes the key strengths of the data source for AI/AN/NA research, as identified by the reviewers.
21. Limitations—summarizes the key limitations of the data source identified by the reviewers.
22. Other—contains other important information identified by the reviewer.

23. Access Requirements and Use Restrictions—describes the steps and cost involved in accessing, if possible, the data set.
24. Contact Information—typically provides a name, telephone number, address, and email for the office responsible for the data collection. However, in some cases only a name is provided, and in other cases, only an email or Internet link is provided.
25. Reports of Interest—provides a non-exhaustive list of key reports related to the data source identified in the course of the project.

2.2 Methodology

The approach used to compile the data catalog consisted of four steps: initial listings of data sources, screening of data sources, reviewing of the data sources, and developing the catalog profiles. Each of these steps is described in more detail in the paragraphs that follow.

Developing the initial listings of data sources. The project team (in consultation with ASPE, our project consultants, and our small AI/AN/NA workgroup) initially developed a substantial list of potential data sources drawing from the following sources:

- Web sites of all federal departments (e.g., Departments of Justice, Defense, Agriculture) with information on publicly available data sets and on administrative data sets that could be made available from each of these departments;
- Data repositories including National Library of Medicine, the Interuniversity Consortium for Political and Social Research (ICPSR) at the University of Michigan, and the University of Virginia to identify additional data sources located in these repositories that may be appropriate;
- Web site developed by the DHHS Data Council that provides a *Directory of Health and Human Services Data Resources* listing available DHHS data resources (http://aspe.hhs.gov/datacncl/datadir);
- Web site developed by DHHS' Office of Minority Health (OMH) and the Office of the Assistance Secretary for Planning and Evaluation (ASPE) under the auspices of the DHHS Data Council entitled *Health and Human Services Statistics About Minorities* (http://www.hhs-stat.net/omh/index.htm);
- Consultation with members of the DHHS Data Council's Racial and Ethnic Data Working Group who served as a workgroup for this project; and
- Consultation with project consultants and a small AI/AN/NA workgroup representing the American Indian and Native Hawaiian communities.

As the project progressed, data sources were continually added to the review list as they were discovered or recommended to the project team. In all, a total of 152 possible data sources were considered.

Screening of data sources. Data sources were initially screened by reviewers to determine their appropriateness for inclusion in the catalog. The screening protocol comprised the first few pages of the review protocol located in Appendix C to this document. This screening protocol ensured that the data sources to be fully profiled and considered for inclusion in the catalog met the technical parameters listed in Section 2 above. Specifically, the reviewers provided the following information on each data source screened:

- Name of data source being reviewed;
- Sponsoring agency;
- Policy issue relevant to data source;
- Brief description of data set;
- Whether the data source identified people who are AI/AN/NA;
- Availability of data set to researchers;
- Cost associated with use of the data set;
- Relevant contact and location information regarding this data source;
- Source of funding for data set;
- Total unweighted AI/AN/NA population in data source and if unweighted n < 100, feasibility of combining across multiple iterations of data collection to increase the size of the population;
- Geographic coverage of the data set;
- Timeframe of the data set;
- Nature of data source (i.e., survey, program reporting data, or registry) and availability of relevant documentation; and
- Qualitative or quantitative nature of data.

Based on this information, the project team and ASPE project officers determined whether the data source should be fully profiled and likely included in the catalog, included in the list of non-reviewed data sources (Section 6), or omitted entirely.

Reviewing the data sources. Data sources appropriate for profiling were then researched and the data were collected about as many of the 25 elements as possible by the initial reviewers. (See Section 2.1 above.)

For some data sources, not all of this information was available through the Internet, published materials, or direct contacts with employees at the agency or data collection facility. After a reasonable effort, reviewers were instructed to note that missing information could not be obtained or was unavailable. Where the information on any of these categories is not applicable to the data set, the category will not appear in the profile (e.g., weighting information for registries) but when information important to the assessment of the utility of the data source was not available, the category remains in the profile and is duly noted.

Initial reviews were evaluated by senior project staff and revisions were made in light of their comments and questions. These revised reviews were then converted to the catalog entries presented here. In summary, a total of 152 possible data sources were considered. Due to resource issues and some early elimination of unusable data sources from the review process, 110 of the 152 data sources were screened. From the 110 screened data sources, 68[6] data sources met the criteria for inclusion and are fully profiled in the catalog.

[6] One data source was not originally considered for inclusion, but was added to the catalog after an agency sent in a full profile.

After all data reviews were completed, the data source profiles, where possible, were forwarded to the agency or point of contact for that data source for review. Westat received comments or approval on 76 percent of the data profiles in this catalog. After agencies and/or points of contact completed their review of the data source profiles, we revised many of the profiles in response to their comments. Westat then submitted the revised data catalog to members of the project DHHS workgroup, the project AI/AN/NA workgroup, the five Westat project consultants, and senior DHHS staff for review. Many of their comments and suggestions were incorporated into the final catalog. A list of the data sets by the different sponsoring agencies is included in Appendix D.

2.3 Using These Data Sources

The data sources described in this catalog fall into four broad categories: publicly available data sets, restricted use data sets, published tables (with or without on-line tabulation capability), and published tables with special tabulations available. Each of these data source types is described briefly below.

Publicly available data sets are collections of raw data records, usually in numeric format that can be downloaded or transmitted by disk, CD, or email to users for analysis. These may contain one record per person, per interaction, or per household. The researcher can access the data in these data sets in order to seek answers to specific questions he/she has. However, analysis of these data sets requires the appropriate computer equipment, a data analysis program (e.g., SAS, SPSS), and a data analyst skilled in handling raw numeric data. Additionally, analysis of some datasets from research efforts that used complex sampling procedures to select respondents may require the use of software designed to correctly estimate variances, such as WesVar or SUDAAN. These data are termed "publicly available" because all identifying information has been removed to allow their use by researchers other than those who collected the data. Publicly available data may be free to users or there may be a fee for acquiring the data. In some cases, potential users may have to complete a data access form and confidentiality agreement. Some examples of publicly available data sets included in this catalog are the American Housing Survey, the National Health Interview Survey, and the Adoption and Foster Care Analysis and Reporting System.

Restricted use data sets are also collections of raw data records that researchers can analyze in a way similar to the publicly available data sets. They also require computer equipment and special software to analyze. They differ from publicly available data sets in that their use is carefully monitored and restricted. Usually, these restrictions are in place to protect confidential information stored in the data set (e.g., names, addresses, income). Holders of restricted use data sets always require potential secondary users to complete a data access agreement. Some data sources require that potential users also submit a proposal for how they will use the data (e.g., research questions to be addressed), proof of financial support for the analyses, and personal information on the users. Some sources may also require that the users perform their analyses in a designated location. There may or may not be a charge for access to these data sets. Some examples of restricted use data sets included in this catalog are the National Survey of Family Growth and the American Community Survey (full sample).[7]

[7] Employees of federal agencies and departments may have access to some restricted use data sources that are not available to non-federal employees. These restricted-use data sources may have more identifying information than non-restricted use versions of the same data source. In some cases, they may allow identification of race of the respondent or participant where the public-use files do not (e.g., the National Immunization Survey).

In some cases, data sources that will not permit access to the raw data, or have extremely limited access to the raw data, have developed on-line analytical capability that will allow public users to conduct some analyses on their own. Users may or may not have full access to the data and types of analyses may be restricted. One example of a data source that allows very limited access to the data but has on-line analytical capabilities included in this catalog is the American Community Survey (full sample).

Data sources that supply only published tables will not allow secondary users access to the raw data. Once the data are collected and analyses are completed, large tabular volumes containing the results from the analyses are published. Potential users can report only the information presented in these tables. For this catalog, these data sources are included only if the published tables have been determined to be of use to researchers and policy makers interested in issues related to the health and well-being of American Indians, Alaska Natives, or other Native Americans. Some examples of data sources that supply only published tables included in this catalog are the Pediatric Nutrition Surveillance System and the Pregnancy Nutrition Surveillance System.

Data sources that both supply published tables and conduct special request tabulations typically do not allow users access to the raw data and have large published tabular volumes. They, however, also have staff that can perform additional analyses at user request to address questions not covered in the tabular volume. The complexity of the additional analyses that can be conducted may vary across sources as will the possible charge for these special requests. Two examples of data sources included in this catalog that conduct special request tabulations are the Resource Patient and Management System, which is the program reporting database for the Indian Health Service, and the Census of Agriculture.

3. Data Sources by Topical Area

The table below groups the 68 data sets included in this catalog by the primary identified policy issue related to the data set. The breakdown of the primary policy issues are as follows:

- Child well-being (4 data sets)
- Demographic and economic indicators (3 data sets)
- Economic well-being (6 data sets)
- Education (7 data sets)
- Elder well-being (2 data sets)
- Family well-being (4 data sets)
- Health policy issues (30 data sets)
- Housing issues (3 data sets)
- Justice system issues (7 data sets)
- Military service/Veterans issues (1 data set)
- Transportation (1 data set)

As many of these data sources cover a variety of topics, Table 1 also includes additional policy issues that are relevant to the data set.

Table 1. Data Sources by Policy Issue

Child Well-being	**Additional Relevant Policy Issues**	**Page #**
Adoption and Foster Care Analysis and Reporting System (AFCARS)	Health policy issues Justice system issues	5-4
National Child Abuse and Neglect Data System (NCANDS)	Justice system issues	5-104
National Survey of Child and Adolescent Well-being (NSCAW)	Family well-being Health policy issues	5-142
Runaway and Homeless Youth Management Information System (RHYMIS)	Housing issues	5-177
Demographic and Economic Indicators	**Additional Relevant Policy Issues**	**Page #**
American Community Survey (ACS)	Economic well-being Family well-being Housing issues	5-8
Census 2000	Economic well-being Education Family well-being Health policy issues Housing issues Transportation	5-33
Census 2000 - The American Indian and Alaska Native Summary File	Economic well-being Education Family well-being Health policy issues Housing issues Transportation	5-38

Table 1 continued.

Economic Well-being	**Additional Relevant Policy Issues**	**Page #**
Census of Agriculture (2002)		5-41
Consumer Expenditure Surveys (CE) Interview and Diary Surveys		5-49
Current Population Survey (CPS)	Demographic and Economic Indicators	5-53
Panel Study of Income Dynamics (PSID)	Child well-being Elder well-being Family well-being Housing issues	5-164
Small Area Income and Poverty Estimates (SAIPE)		5-180
Survey of Program Dynamics (SPD)	Child well-being Education Family well-being Health policy issues Transportation Military service	5-188
Education	**Additional Relevant Policy Issues**	**Page #**
Early Childhood Longitudinal Study, Birth Cohort (ECLS-B)	Child well-being Family well-being	5-58
Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K)		5-63
Head Start Program Information Report	Child well-being Family well-being	5-73
Integrated Postsecondary Education Data System (IPEDS)		5-83
National Assessment of Adult Literacy (NAAL)		5-101
National Household Education Surveys Program (NHES)		5-120
National Indian Education Study (NIES)		5-130
Elder Well-being	**Additional Relevant Policy Issues**	**Page #**
Health and Retirement Study (HRS)	Economic well-being Education Health policy issues	5-75
National Aging Program Information Systems (NAPIS) State Performance Reports		5-96
Family Well-being	**Additional Relevant Policy Issues**	**Page #**
Food Stamp Program Quality Control Database (FSPQC)	Economic well-being	5-67
National Survey of America's Families (NSAF)	Child well-being Economic well-being Health policy issues Housing issues	5-137
National Survey of Family Growth (NSFG)	Health policy issues	5-146
Temporary Assistance for Needy Families (TANF) and Tribal TANF	Economic well-being	5-191

Table 1 continued.

Health Policy Issues	Additional Relevant Policy Issues	Page #
Behavioral Risk Factor Surveillance System (BRFSS)		5-22
Consumer Assessment of Healthcare Providers and Systems (CAHPS) Health Plan Survey Response Data		5-25
California Health Interview Survey (CHIS)	Child well-being Elder well-being	5-29
Hawaii Health Survey (HHS)		5-70
Health Behavior in School-aged Children (HBSC)	Economic well-being Education Child well-being Family well-being	5-78
Health Information National Trends Survey (HINTS)		5-81
Medicaid Analytic Extract (MAX)	Child well-being Elder well-being	5-85
Medical Expenditure Panel Survey (MEPS)	Child well-being Elder well-being Family well-being	5-88
Medicare Denominator Files	Elder well-being	5-92
Medicare Utilization – Standard Analytic Files (SAFs)		5-94
National Ambulatory Medical Care Survey (NAMCS)		5-98
National Epidemiologic Survey on Alcohol and Related Conditions (NESARC)		5-111
National Health Interview Survey (NHIS)	Child well-being Elder well-being Family well-being	5-114
National Hospital Ambulatory Medical Care Survey (NHAMCS)		5-117
National Longitudinal Mortality Study (NLMS)		5-133
National Mortality Followback Survey (NMFS)		5-135
National Survey on Drug Use and Health (NSDUH)		5-153
National Vital Statistics System: Linked Birth-Infant Death (NVSS-I)		5-156
National Vital Statistics System: Mortality (NVSS-M)		5-159
National Vital Statistics System: Natality (NVSS-N)		5-162
Pediatric Nutrition Surveillance System (PedNSS)	Child well-being	5-167
Pregnancy Nutrition Surveillance System (PNSS)		5-169
Pregnancy Risk Assessment Monitoring System (PRAMS)	Child well-being	5-171
Resource and Patient Management System (RPMS) and National Patient Information Reporting System (NPIRS)		5-174
Surveillance, Epidemiology, and End Results (SEER)		5-183
Tobacco Use Supplement to the Current Population Survey (TUS-CPS)		5-194
Treatment Episode Data Set (TEDS)		5-199
United States Renal Data System (USRDS)		5-205
Washington State Population Survey (WSPS)	Economic well-being	5-207
Youth Risk Behavior Surveillance Survey (YRBSS)		5-211

Table 1 continued.

Housing Issues	**Additional Relevant Policy Issues**	**Page #**
A Picture of Subsidized Households	Economic well-being	5-2
American Housing Survey (AHS)		5-12
American Housing Survey: Metropolitan Surveys		5-15
Justice System Issues	**Additional Relevant Policy Issues**	**Page #**
Annual Survey of Jails (ASJ)		5-19
Census of Jails		5-44
Census of Tribal Justice Agencies in Indian Country		5-47
National Crime Victimization Survey (NCVS)		5-108
Survey of Jails in Indian Country (SJIC)		5-186
Uniform Crime Reports		5-202
Youth Gangs in Indian Country		5-209
Military Service/Veterans Issues	**Additional Relevant Policy Issues**	**Page #**
National Survey of Veterans (NSV)		5-150
Transportation	**Additional Relevant Policy Issues**	**Page #**
National Household Travel Survey (NHTS)		5-125

4. Data Sources by Subpopulation Coverage

The 68 data sets included in this catalog vary in the way AI/AN/NA individuals are identified. Race is directly self-reported by the respondents or the respondents' legal guardian (if the focal respondent is a minor) in 35 of these studies. Twenty-eight of these studies obtain race information from administrative records (e.g., enrollment forms, patient files). Five of the data sets do not identify AI/AN/NA individuals.

Although there is some variation in the categories used to identify AI/AN/NA individuals, the majority of the studies in this catalog use the broader categories "American Indian or Alaska Native" and "Native Hawaiian or other Pacific Islander." These broad categories meet the minimum racial classification standards set by the Office of Management and Budget (OMB)[8]. Table 2 below lists the data sets by the subpopulation coverage for the American Indian or Alaska Native racial category, while Table 3 lists the data sets by the subpopulation coverage for the Native Hawaiian or other Pacific Islander subcategory.

Table 2. Data Sources by Subpopulation Coverage: American Indian/Alaska Native

Data Sources that identify Tribal Affiliation	**Notes**	**Page #**
American Community Survey (ACS)		5-8
California Health Interview Survey (CHIS)	Identifies 10 federally recognized tribes	5-29
Census 2000		5-33
Census 2000 - The American Indian and Alaska Native Summary File	Identifies 1,081 federally recognized tribes and villages	5-38
Early Childhood Longitudinal Study, Birth Cohort (ECLS-B)		5-58
Pediatric Nutrition Surveillance System (PedNSS)	Identifies 7 federally recognized tribes	5-167
Pregnancy Nutrition Surveillance System (PNSS)	Identifies 6 federally recognized tribes	5-169
Resource and Patient Management System (RPMS) and National Patient Information Reporting System (NPIRS)	Analysis by tribe is dependent on tribal approval	5-174
Data sources that include the category American Indian alone (AI)	**Notes**	**Page #**
American Community Survey (ACS)		5-8
Census 2000		5-33
Health and Retirement Study (HRS)	This category is available beginning in 2006	5-75
Resource and Patient Management System (RPMS) and National Patient Information Reporting System (NPIRS)		5-174
Treatment Episode Data Set (TEDS)		5-199

[8] www.whitehouse.gov.omb/fedreg/1997standards

Table 2 continued.

Data sources that include the category Alaska Native alone (AN)	**Notes**	**Page #**
American Community Survey (ACS)		5-8
Census 2000		5-33
Health and Retirement Study (HRS)	This category is available beginning in 2006	5-75
Resource and Patient Management System (RPMS) and National Patient Information Reporting System (NPIRS)		5-174
Treatment Episode Data Set (TEDS)		5-199
Data sources that include the combined category American Indian/Alaska Native (AI/AN)	**Notes**	**Page #**
A Picture of Subsidized Households	Category is "Native American"	5-2
Adoption and Foster Care Analysis and Reporting System (AFCARS)		5-4
American Housing Survey		5-12
American Housing Survey: Metropolitan Surveys		5-15
Annual Survey of Jails (ASJ)		5-19
Behavioral Risk Factor Surveillance System (BRFSS)		5-22
CAHPS Health Plan Survey Response Data		5-25
California Health Interview Survey (CHIS)		5-29
Census of Agriculture (2002)		5-41
Census of Jails		5-44
Consumer Expenditure Surveys (CE) Interview and Diary Surveys		5-49
Current Population Survey (CPS)		5-53
Early Childhood Longitudinal Study, Birth Cohort (ECLS-B)		5-58
Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K)		5-63
Food Stamp Program Quality Control Database (FSPQC)		5-67
Head Start Program Information Report		5-73
Health Behavior in School-aged Children (HBSC)		5-78
Health Information National Trends Survey (HINTS)		5-81
Integrated Postsecondary Education Data System (IPEDS)		5-83
Medicaid Analytic Extract (MAX)		5-85
Medical Expenditure Panel Survey (MEPS)		5-88
Medicare Denominator Files	Category is "North American Native"	5-92
Medicare Utilization – Standard Analytic Files (SAFs)	Category is "North American Native"	5-94
National Aging Program Information Systems (NAPIS) State Performance Reports		5-96
National Ambulatory Medical Care Survey (NAMCS)		5-98
National Assessment of Adult Literacy (NAAL)		5-101
National Child Abuse and Neglect Data System (NCANDS)		5-104
National Crime Victimization Survey (NCVS)		5-108
National Epidemiologic Survey on Alcohol and Related Conditions (NESARC)		5-111
National Health Interview Survey (NHIS)		5-114

Table 2 continued.

Data sources that include the combined category American Indian/Alaska Native (AI/AN) - Continued	Notes	Page #
National Hospital Ambulatory Medical Care Survey (NHAMCS)		5-117
National Household Education Surveys Program (NHES)		5-120
National Household Travel Survey (NHTS)		5-125
National Indian Education Study (NIES)		5-130
National Longitudinal Mortality Study (NLMS)		5-133
National Mortality Followback Survey (NMFS)	Category is "American Indian (AI), Aleut or Eskimo"	5-135
National Survey of America's Families (NSAF)	Category is "American Indian (AI), Aleut or Eskimo"	5-137
National Survey of Child and Adolescent Well-being (NSCAW)		5-142
National Survey of Family Growth (NSFG)		5-146
National Survey of Veterans (NSV)		5-150
National Survey on Drug Use and Health (NSDUH)		5-153
National Vital Statistics System: Linked Birth-Infant Death (NVSS-I)		5-156
National Vital Statistics System: Mortality (NVSS-M)		5-159
National Vital Statistics System: Natality (NVSS-N)		5-162
Panel Study of Income Dynamics (PSID)	Category is "Native American"	5-164
Pediatric Nutrition Surveillance System (PedNSS)		5-167
Pregnancy Nutrition Surveillance System (PNSS)		5-169
Pregnancy Risk Assessment Monitoring System (PRAMS)		5-171
Runaway and Homeless Youth Management Information System (RHYMIS)		5-177
Surveillance, Epidemiology, and End Results (SEER)		5-183
Survey of Program Dynamics (SPD)	Category is "American Indian (AI), Aleut or Eskimo"	5-188
Temporary Assistance for Needy Families (TANF) and Tribal TANF		5-191
Tobacco Use Supplement to the Current Population Survey (TUS-CPS)		5-194
Uniform Crime Reports		5-202
United States Renal Data System (USRDS)	Category is "Native American"	5-205
Washington State Population Survey (WSPS)		5-207
Youth Risk Behavior Surveillance Survey (YRBSS)		5-211

Table 3. Data Sources by Subpopulation Coverage: Native Hawaiian/Other Pacific Islander

Data sources that include the category Native Hawaiian alone (NH)	**Notes**	**Page #**
American Community Survey (ACS)		5-8
Census 2000		5-33
Early Childhood Longitudinal Study, Birth Cohort (ECLS-B)		5-58
Hawaii Health Survey (HHS)		5-70
Health and Retirement Study (HRS)	This category is available beginning in 2006	5-75
National Survey of Veterans (NSV)		5-150
National Vital Statistics System: Linked Birth-Infant Death (NVSS-I)	Category is not available after 2003	5-156
National Vital Statistics System: Mortality (NVSS-M)		5-159
National Vital Statistics System: Natality (NVSS-N)	Category is not available after 2003	5-162
Pregnancy Risk Assessment Monitoring System (PRAMS)		5-171
Data sources that include the category Other Pacific Islander alone (PI)	**Notes**	**Page #**
American Community Survey (ACS)		5-8
Census 2000		5-33
Early Childhood Longitudinal Study, Birth Cohort (ECLS-B)		5-58
Health and Retirement Study (HRS)	This category is available beginning in 2006	5-75
National Survey of Veterans (NSV)		5-150
National Vital Statistics System: Mortality (NVSS-M)	Includes categories for Samoan and Guamanian	5-159
Data sources that include the combined category Native Hawaiian or other Pacific Islander (NH/PI)	**Notes**	**Page #**
Adoption and Foster Care Analysis and Reporting System (AFCARS)		5-4
American Housing Survey		5-12
American Housing Survey: Metropolitan Surveys		5-15
Annual Survey of Jails (ASJ)		5-19
Behavioral Risk Factor Surveillance System (BRFSS)		5-22
CAHPS Health Plan Survey Response Data		5-25
California Health Interview Survey (CHIS)		5-29
Census of Agriculture (2002)		5-41
Census of Jails		5-44
Consumer Expenditure Surveys (CE) Interview and Diary Surveys		5-49
Current Population Survey (CPS)		5-53
Early Childhood Longitudinal Survey Kindergarten Cohort (ECLS-K)		5-63
Food Stamp Quality Control Database (FSPQC)	Category available beginning in 2007	5-67
Head Start Program Information Report		5-73
Health Behavior in School-aged Children (HBSC)		5-78
Health Information National Trends Survey (HINTS)		5-81
Medicaid Analytic Extract (MAX)		5-85

Table 3 continued.

Data sources that include the combined category Native Hawaiian or other Pacific Islander (NH/PI) - Continued	**Notes**	**Page #**
Medical Expenditure Panel Survey (MEPS)		5-88
National Ambulatory Medical Care Survey (NAMCS)		5-98
National Assessment of Adult Literacy (NAAL)		5-101
National Child Abuse and Neglect Data System (NCANDS)		5-104
National Crime Victimization Survey (NCVS)	Category available beginning in 2003	5-108
National Epidemiologic Survey on Alcohol and Related Conditions (NESARC)		5-111
National Hospital Ambulatory Medical Care Survey (NHAMCS)		5-117
National Household Education Surveys (NHES)	Category available beginning in 2005	5-120
National Household Travel Survey (NHTS)		5-125
National Longitudinal Mortality Study (NLMS)		5-133
National Survey of Family Growth (NSFG)		5-146
National Survey on Drug Use and Health (NSDUH)		5-153
Runaway and Homeless Youth Management Information System (RHYMIS)		5-177
Temporary Assistance for Needy Families (TANF) and Tribal TANF		5-191
Tobacco Use Supplement to the Current Population Survey (TUS-CPS)		5-194
Washington State Population Survey (WSPS)		5-207
Youth Risk Behavior Surveillance Survey (YRBSS)		5-211

5. Listing of Data Sources

The following pages contain detailed descriptions of the 68 data sources included in this catalog.

A Picture of Subsidized Households (1998)

Sponsor: U.S. Department of Housing and Urban Development (HUD)

Description: The Picture of Subsidized Households (1998) presents summary information reported by subsidized housing programs across the nation. The program data included in the data set comes from five different sources:

1. Indian Housing
2. Public Housing
3. Section 8 housing (including Section 8 Certificates and Vouchers, Section 8 Moderate Rehabilitation, and Section 8 New Construction or Substantial Rehabilitation)
4. Federal Housing Administration (FHA) (including Section 236 projects and other FHA projects with subsidy such as Section 8 Loan Management, Rental Assistance Program (RAP), Rent Supplement, Property Disposition, etc.)
5. Low Income Housing Tax Credit

The data file and reports cover about four and a half million HUD-subsidized housing units, and a third of a million housing units assisted by Low Income Housing Tax Credits, for a total of nearly five million subsidized housing units.

Relevant Policy Issues: Income Status, Economic Assistance Program Participation Rates, and Housing Quality.

Data Type(s): Program reporting data

Unit of Analysis: Housing subsidy program

Identification of AI/AN/NA: The 1998 data set includes the percent of Native Americans (defined as American Indians and Alaska Natives) living in subsidized housing in 1998.

AI/AN/NA Population in Data Set: This data set is summary data rather than raw data, and as such, does not provide counts of AI/AN/NA persons but does provide the percentage of Native Americans living in subsidized housing. Of all those living in subsidized housing in 1998, across all types of subsidized housing, 1 percent were Native American. Of those living in Indian housing, 89 percent were Native American.

Geographic Scope: The geographic scope of the study is national. Geographic indicators available in the data include latitude, longitude, zip code, county, MSA, census tract, and Congressional district.

Date or Frequency: The data used in this report and compiled into the data set are collected continuously and only periodically compiled into the report and data sets. *A Picture of Subsidized Households* previously was produced the 1970s and in 1996, and 1997. *A Picture of Subsidized Households: 2000* and *A Picture of Subsidized Households: 2004* are forthcoming (both are expected by the end of

the 2006 calendar year). These reports will not contain any data on Indian Housing but will include information on American Indians, Alaska Natives, and other Native Americans living in non-Indian Housing. Tentative plans to release *A Picture of Subsidized Households* reports biennially are under consideration.

Data Collection Methodology: Subsidized housing programs submit reports to HUD.

Participation: Mandatory

Strengths: Data are collected on key policy issues regarding housing. There are multiple years of data available.

Limitations: These data do not include any information on Native Hawaiians or other Pacific Islanders living in subsidized housing.

Access Requirements and Use Restrictions: Data set is available to public at no cost.

Contact Information: General user information can be accessed through HUD USER:

HUD USER
P.O. Box 23268
Washington, DC 20026-3268
Toll Free: (800) 245-2691
TDD: (800) 927-7589
Local: (202) 708-3178
Fax: (202) 708-9981
email: helpdesk@huduser.org

More detailed information can be accessed by contacting:
Robert W. Gray
U.S. Department of Housing and Urban Development
451 7th Street S.W., Washington, DC 20410
Robert_W._Gray@hud.gov

The 1998 data and documentation are available for download from the following website: http://www.huduser.org/datasets/assthsg/statedata98/.

Adoption and Foster Care Analysis and Reporting System (AFCARS)

Sponsor: U.S. Department of Health and Human Services (DHHS)/Administration for Children and Families (ACF)

Description: The Adoption and Foster Care Analysis and Reporting System (AFCARS) provides child-specific information on all children covered by the protections of Title IV-B and Title IV-E of the Social Security Act. On a semi-annual basis, all states submit data to the U.S. Children's Bureau concerning each child in foster care and each child who has been adopted under the authority of the state's child welfare agency. The AFCARS databases have been designed to address policy development and program management issues at both the state and federal levels. The data are also useful for researchers interested in analyzing aspects of the United States' foster care and adoption programs.

For each year since 1995, there are two AFCARS files, one containing adoption data and the other containing foster care data. These annual files are constructed from the states' semi-annual data submissions. The adoption file contains 45 data elements concerning the adopted child's gender, race, birth date, ethnicity and prior relationship with the adoptive parents. The date the adoption was finalized, dates parental rights were terminated, characteristics of birth and adoptive parents, and whether the child was placed from within the United States or from another country are also captured. The foster care file contains 89 elements providing information on child demographics including gender, birth date, race, and ethnicity. Information about the number of previous stays in foster care, service goals, availability for adoption, dates of removal and discharge, funding sources, and the biological and foster parents is also included in the foster care files.

Relevant Policy Issues: Measurement of Health Status, Measures of Well-being for Children, Factors Contributing to Well-being Disparities of Children, and Child Maltreatment Rates.

Data Type(s): Registry

Unit of Analysis: Individual (child)

Identification of AI/AN/NA: Instructions for states reporting race are: "In general, a person's race is determined by how others define them or by how they define themselves. In the case of young children, parents determine the race of the child. Data entry staff are to indicate all races that apply."

<u>Definitions of racial categories include:</u>
American Indian or Alaska Native (AI/AN) is defined as a person having origins in any of the original peoples of North America or South America (including Central America), and who maintains tribal affiliation or community attachment.

Native Hawaiian or other Pacific Islander (NH/PI) is defined as a person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.

Asian – A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.

Black or African American – A person having origins in any of the black racial groups of Africa.

White – A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.

AI/AN/NA Population in Data Set:

For FY 2003, number of children in foster care:
TOTAL: 523,000
AI/AN non-Hispanic: 10,260
NH/PI non-Hispanic: 1,540

For FY 2003, number of children waiting to be adopted:
TOTAL: 119,000
AI/AN non-Hispanic: 2,190
NH/PI non-Hispanic: 340

For FY 2003, number of children who were adopted:
TOTAL: 50,000
AI/AN non-Hispanic: 700
NH/PI non-Hispanic: 130

Geographic Scope:

The geographic scope of the AFCARS is national. A state indicator is available for all records on both the adoption and foster care files. For foster care data, the geographic Federal Information Processing Standards (FIPS) codes for the local county agency responsible for the case are available for all children in counties with more than 1,000 records. (If county has less than 1,000 records, FIPS codes are excluded for reasons of confidentiality.) Limited geographic analysis is possible by state and county.

Date or Frequency:

States are required to submit AFCARS data semi-annually to ACF. The AFCARS reporting periods are October 1 through March 31 and April 1 through September 30. Data for each reporting period are due no later than May 15 and November 14, respectively. An annual file is constructed from the two semi-annual files.

Data collection has been ongoing since 1995, although use of the data prior to 1998 is discouraged. The period from 1995 to 1997 was a start-up phase for AFCARS. Many states were still developing their information systems and were unable to submit data. Other states were able to submit data, but the quality was

	either missing or poor on many of the data elements. Therefore, pre-1998 data sets are not as complete or reliable as the data for subsequent years. Since 1998, participation by the states has been universal and the data quality has improved dramatically. The most recent data available are for FY2004.
Data Collection Methodology:	States submit case-level reports electronically to AFCARS.
Participation:	Mandatory
Response Rate:	By 1999, all states (plus the District of Columbia (DC) and Puerto Rico) had submitted adoption data. By 2001, all states, DC and Puerto Rico had submitted foster care data.
Authorization:	Under the final AFCARS' rule, states are required to collect data on all adopted children covered by Title IV-B/E of the Social Security Act.
Strengths:	AFCARS data sets contain a large number of AI/AN/NA respondents. The data are collected on key policy issues, including child welfare. There are multiple years of data available. The documentation is thorough and clear. Mandatory reporting means high compliance among states.
Limitations:	While data collection began in 1995, use of the data prior to 1998 is discouraged. Tribal agencies who place children for adoption voluntarily report data to AFCARS. Since these adoptions do not involve a state agency, these records are not included in the publicly available version of the data. A 2003 report issued by the Department of Health and Human Services and entitled *Adoption and Foster Care Analysis and Reporting System (AFCARS): Challenges and Limitations* reports there is some inconsistency in the definition of data elements, most notably problems with placement and date-of-discharge definitions.
Other:	In 2000, technical changes to the race/ethnicity data elements in AFCARS were made. Previously, race categories were mutually exclusive, but starting in 2000 children could be classified as multi-racial.
Access Requirements and Use Restrictions:	Researchers need to complete and submit a Terms of Use Agreement.

Adoption and Foster Care Analysis and Reporting System (AFCARS)
(continued)

Contact Information:	The data set is available through the National Data Archive on Child Abuse and Neglect at Cornell University: http://www.ndacan.cornell.edu. National Data Archive on Child Abuse and Neglect Beebe Hall - FLDC Cornell University Ithaca NY 14853 Phone: (607) 255-7799 Fax: (607) 255-8562 E-Mail: NDACAN@cornell.edu

American Community Survey (ACS)

Sponsor: U.S. Department of Commerce/U.S. Census Bureau

Description: The American Community Survey (ACS) was designed to provide current estimates of community change, and intended to replace the decennial Census long form by collecting and producing updated population and housing information every year instead of every 10 years. About three million households are to be surveyed each year. The ACS collects information from U.S. households similar to what was collected on the Census 2000 long form, such as income and employment, commute time to work, home value and expenses, type of housing, household composition, health status, and veteran status. The ACS began testing in 1996 and expanded to a national demonstration design from 2000 through 2004. Full implementation into all counties began in 2005.

Each year, a subsample of the ACS is selected to construct the Public Use Microdata (PUMS). This data set is available online for all researchers. Use of the full sample of the ACS is restricted and access is difficult to obtain. However, the full sample data are used to create the published population estimates from the ACS. Both the PUMS data and the restricted-access full ACS sample are discussed in this profile.

Relevant Policy Issues: Demographic and Economic Indicators, Income Disparities, Unemployment Rates, Economic Assistance Program Participation Rates, Measures of Well-being for Families/Households, Housing Quality, Type of Housing, Housing Ownership, and Rental Unit Quality and Cost.

Data Type(s): Survey

Unit of Analysis: Individual and Household.

Identification of AI/AN/NA: Race is self-reported. One household member fills out the ACS questionnaire and reports race for himself/herself and all other members of the household. Instructions in the ACS survey for reporting race are: "What is this person's race? Mark one or more races to indicate what this person considered himself/herself to be." Response categories include:

- White
- Black or African American
- American Indian or Alaska Native – Print name of enrolled or principal tribe (AI/AN)
- Asian Indian
- Chinese
- Filipino
- Japanese
- Korean
- Vietnamese
- Other Asian – Print race

- Native Hawaiian (NH)
- Guamanian or Chamorro
- Samoan
- Other Pacific Islander – Print race below (OPI)
- Some other race – Print race below

Additionally, the questionnaire contains several questions that either require or permit respondents to write-in their responses. The write-in fields cover the following topics: race, Hispanic origin, place of birth, ancestry, migration, language, place of work, industry and occupation. These write-in responses are then coded. Using these coded responses, further race classification of the data is possible, such as separating "American Indian" from "Alaska Native."

AI/AN/NA Population in Data Set:

As noted earlier, there is both the restricted-access full sample of the ACS and the Public Use Microdata set (PUMS). The restricted-access sample provides race categories in greater detail than the PUMS data, while the race categories for the PUMS data are combined into fewer categories.

At the time of this publication, the data from the 2005 full implementation of the ACS are not yet available. The PUMS data will contain about 40 percent of the full ACS sample and represent 1 percent of the population. Based on information from the pilot testing, the PUMS data should contain an adequate number of AI/AN/NA individuals to support analysis targeting this population.

AI/AN/NA Subpopulations:

The ACS questionnaire asks for AI/AN individuals to give the name of their enrolled or principal tribe. Total population estimates are provided for the following tribal groupings: Cherokee, Chippewa, Navajo and Sioux. Tribal affiliation is not available in the PUMS data set, but is present in the full restricted ACS data.

The Native Hawaiian/Other Pacific Islander race category is broken out into the following subcategories and total population estimates are reported separately: Native Hawaiian, Guamanian or Chamorro, Samoan, Other Pacific Islander. Again, this level of detail is only available in the restricted use ACS data. The PUMS data only reports the combined category Native Hawaiian/Other Pacific Islander.

Geographic Scope:

The geographic scope of the ACS is national. The ACS identifies the following individual geographic areas: nation, state, county, county subdivision, place-county, place, metropolitan statistical area (MSA)/consolidated metro statistical area (CMSA)/primary metropolitan statistical area (PMSA), and Congressional district. The list of geographic identifiers will expand with the release of the 2005 data.

Data from the 2004 ACS are available for over 800 geographic areas, including 244 counties, 203 Congressional districts, most metropolitan areas of 250,000 population or more, all 50 states, and the District of Columbia. From the mid

1990s to 2004, the ACS survey was administered to selected sites. The ACS data collection effort was fully implemented in 2005. The Census Bureau plans to begin releasing the 2005 survey estimates in the summer of 2006, but only for areas with populations of 65,000 or more. For areas with populations of 20,000 or more, data release is planned for the summer of 2008, with estimates for all areas–down to census tract/block group level–by the summer of 2010.

Date or Frequency: Data collection is conducted on a monthly basis. Results are compiled and published annually. At the time of this reporting, the most current available data are for 2004. Release dates are noted in the Geographic Scope section above.

Data Collection Methodology: Surveys are mailed every month to a random sample of addresses in each county. The self-enumeration procedure uses several mailing pieces, including a prenotification letter, the American Community Survey questionnaire, a reminder card, and a replacement questionnaire if the original questionnaire is not returned in a timely manner. If a household does not respond in 6 weeks, Census Bureau staff will attempt to contact the respondent by telephone to complete the survey. If that, too, fails, about one in every three addresses remaining will be visited by Census Bureau staff for an in-person interview.

Participation: Mandatory

Response Rate: The ACS does not report an unweighted response rate. Weighting is used because not all housing units have the same probability of selection. The weighted response rate for the 2004 ACS is reported as 93.1 percent, due to a special budget issue. From 2001 to 2003, response rates ranged from 96.7 percent to 97.7 percent.

Sampling Methodology: The 2004 ACS used a two-stage stratified annual sample of approximately 830,000 housing units. Step one of the sampling design was to divide the U.S. into primary sampling units (PSUs); step two was grouping the PSUs into strata based on independent information. After that, one pair of PSUs from each stratum was selected. The 2004 population estimates were derived from 568,966 final interviews. Beginning in 2005, the first stage of sample was no longer used, as all counties will be included in the sample.

Oversample of AI/AN/NA Population: A larger proportion of addresses are sampled for small governmental units including American Indian reservations. The monthly sample size is designed to approximate the sampling ratio of Census 2000 Long Form, including the oversampling of small governmental units.

Analysis: The ACS is a weighted data set. The PUMS methodology report gives a detailed explanation on how to apply appropriate weights when conducting analysis and how to calculate standard errors for the survey variable(s) of interest. Each variable in the ACS data set has an associated design effect. The design effects for the race categories in the ACS are listed as:

- White alone = 2.5
- Black or African American alone = 3.1
- AI/AN alone, Asian alone, NH/PI alone, some other race alone = 3.0

Authorization: The American Community Survey is part of the 10-year Census. As such, its legal authority derives from the same statutes that authorize the Census: Title 13 of the U.S. Code, Sections 141 and 193.

Strengths: The data set contains a large number of AI/AN/NA respondents. The data are collected on key policy issues, including housing and economic well-being. There are multiple years of data available.

The AI/AN/NA population is oversampled in the ACS, thereby improving the reliability of estimates for this population by reducing the variance. Coverage rates for the 2004 ACS for AI/AN are reported as 100 percent and as 90 percent for Native Hawaiian and other Pacific Islanders.

Limitations: The tabulations prepared from the PUMS are based on a subset of the 2004 American Community Survey (ACS) sample. Estimates from the ACS PUMS file are expected to be different from the previously released ACS estimates because they are subject to additional sampling error and further data processing operations. The full data set is difficult to obtain (see below).

Access Requirements and Use Restrictions: The ACS PUMS data are available for download at no cost at the Census Bureau's American FactFinder website: http://factfinder.census.gov/home/saff/main.html?_lang=en.

It may be possible for researchers to access the full ACS survey data set, but obtaining permission for this is difficult. Interested researchers can send an application to the Center of Economic Studies. If approved, researchers will need to work with the data at a Census Bureau data site, or at research centers located in various cities.

Contact Information: For general information about the scope and content of the American Community Survey, call 1 (888) 456-7215 or email cmo.acs@census.gov.

Additional information can be found online at: http://www.census.gov/acs/www/index.html

American Housing Survey (2003)

Sponsor: U.S. Department of Housing and Urban Development (HUD)

Description: The American Housing Survey (AHS) collects data on the Nation's housing, including apartments, single-family homes, mobile homes, vacant housing units, household characteristics, income, housing and neighborhood quality, housing costs, equipment and fuels, size of housing unit, and recent movers. The AHS is conducted by field representatives who obtain information from occupants of homes or from informed people such as landlords, rental agents, or knowledgeable neighbors about vacant homes. Interviewing occurs from May 30 through September 8 and is conducted every other year. The 2003 national survey is a sample of about 61,050 designated housing units.

Relevant Policy Issues: Housing Quality, Type of Housing, Housing Ownership, and Rental Unit Quality and Cost.

Data Type(s): Survey

Unit of Analysis: Household. A household consists of all people who occupy a particular housing unit as their usual residence, or who live there at the time of the interview and have no usual residence elsewhere. The usual residence is the place where the person lives and sleeps most of the time. This place is not necessarily the same as a legal residence, voting residence, or domicile. Households include not only occupants related to the householder but also any lodgers, roomers, boarders, partners, wards, foster children, and resident employees who share the living quarters of the householder. It includes people temporarily away for reasons such as visiting, traveling in connection with their jobs, attending school, in general hospitals, and in other temporary relocations. By definition, the count of households is the same as the count of occupied housing units.

Identification of AI/AN/NA: Race is self-reported. Interviewees were asked to respond to the question on race by indicating one or more of six race categories. The six race categories included:

- White
- Black or African American
- American Indian or Alaska Native (AI/AN)
- Asian
- Native Hawaiian or other Pacific Islander (NH/PI)
- Some Other Race (this category is not read or displayed to the respondent)

AI/AN/NA Population in Data Set: The total number of completed surveys for the AHS 2003 was 71,170. Responses to the race item were recoded into the multiple race categories. The following categories reflect the unweighted counts for the AI/AN/NA respondents in the 2003 AHS:

AI/AN Only: 300
NH/PI Only: 139
White/AI/AN: 291
White/NH/PI: 11
Black/AI/AN: 49
Black/NH/PI: 2
AI/AN/Asian: 3
Asian/NH/PI: 10
White/Black/AI/AN: 27
White/Asian/NH/PI: 5
White/AI/AN/Asian: 1

Geographic Scope: The geographic scope of the AHS is national. Geographic analysis is possible by county.

Date or Frequency: From 1973 to 1981, the AHS collected national data every year and was called the Annual Housing Survey. Since 1981, the AHS has been conducted biannually in odd-numbered years.

Data Collection Methodology: In-person interviews and telephone interviews are conducted by field interviewers.

Participation: Optional, without incentives

Response Rate: For 2003, the unweighted overall response rate was 91 percent.

Sampling Methodology: The sample for AHS is spread over 394 primary sampling units (PSUs), counties or groups of counties or independent cities. These PSUs include 878 counties and independent cities with coverage in all 50 states and the District of Columbia. If there were over 100,000 housing units in a PSU at the time of selection, the PSU is known as a self-representing PSU because it was removed from the probability sampling operation and was in the sample with certainty. There are 170 self-representing PSUs. The Census Bureau grouped the remaining PSUs and selected one PSU per group, proportional to the number of housing units in the PSU, to represent all PSUs in the group. These selected PSUs are referred to as nonself-representing PSUs. After this, a sample of housing units was chosen within the selected PSUs.

Authorization: Title 12, Sections 1701Z-1 and 1701Z-2g of the U.S. Code authorize the Secretary of HUD to collect data from public and private agencies and protect the confidentiality of the data. Title 12, Section 1701Z-10 mandates the collection of the data for the AHS.

Strengths: Data are collected on key policy issues, including issues related to housing. There are multiple years of data available.

Limitations: No major limitations were identified.

American Housing Survey (2003)
(continued)

Access Requirements and Use Restrictions:	Data are available to public at no cost.
Contact Information:	The American Housing Survey Branch can be contacted by email at ahsn@census.gov or by phone at (301) 763-3235. Tables, reports and the actual data can be downloaded from http://www.census.gov/hhes/www/housing/ahs/nationaldata.html.

American Housing Survey: Metropolitan Surveys

Sponsor: U.S. Department of Housing and Urban Development (HUD)

Description: The Metropolitan Area Surveys portion of the American Housing Survey (AHS Metro Survey) collects data on housing, including apartments, single-family homes, mobile homes, vacant housing units, household characteristics, income, housing and neighborhood quality, housing costs, equipment and fuels, size of housing unit, and recent movers in a selection of metropolitan areas across the country. Data are gathered for about 14 metropolitan areas in even-numbered years until a total of 47 metropolitan areas have been covered. That is, householders in selected areas are interviewed every 6 years until all 47 metropolitan areas have been surveyed. The cycle begins again 6 years later. Since 1984, each metropolitan area has been represented by a sample of at least 3,200 designated housing units. The units are divided between the central city and the rest of the metropolitan area.

Relevant Policy Issues: Housing Quality, Type of Housing, and Housing Ownership.

Data Type(s): Survey

Unit of Analysis: Individuals, Households, and Metropolitan areas.

A household consists of all people who occupy a particular housing unit as their usual residence, or who live there at the time of the interview and have no usual residence elsewhere. The usual residence is the place where the person lives and sleeps most of the time. This place is not necessarily the same as a legal residence, voting residence, or domicile. Households include not only occupants related to the householder but also any lodgers, roomers, boarders, partners, wards, foster children, and resident employees who share the living quarters of the householder. It includes people temporarily away for reasons such as visiting, traveling in connection with their jobs, attending school, in general hospitals, and in other temporary locations. By definition, the count of households is the same as the count of occupied housing units.

Identification of AI/AN/NA: Race is self-reported. Participants were asked to respond to the question on race by indicating one or more of six race categories. The six race categories were:

- White
- Black or African American
- American Indian or Alaska Native (AI/AN)
- Asian
- Native Hawaiian or other Pacific Islander (NH/PI)
- Some Other Race (this category is not read or displayed to the respondent)

AI/AN/NA Population in Data Set:

The following categories reflect the unweighted counts for the recoded AI/AN/NA relevant response categories in the 2004 AHS Metro Survey:

TOTAL: 62,005
AI/AN: 350
NH/PI: 122
White/AI/AN: 555
White/NH/PI: 27
Black/AI/AN: 74
Black/NH/PI: 6
AI/AN/Asian: 1
Asian/NH/PI: 11
White/Black/AI, AN: 40
White/AI/AN/Asian: 1
White/Asian/NH/PI: 1
White/AI/AN/Asian: 5

Geographic Scope:

The geographic scope of the AHS Metro Survey is national. Geographic analysis is possible by Metropolitan Statistical Area (MSA) in the following groupings or in a combined data set:

The 2004 AHS Metro Survey covered 13 metropolitan areas: Atlanta, Cleveland, Denver, Hartford, Indianapolis, Memphis, New Orleans, Oklahoma, Pittsburgh, St. Louis, Sacramento, San Antonio, Seattle-Everett.

The 2002 AHS Metro Survey covered 13 metropolitan areas: Anaheim-Santa Ana, Buffalo, Charlotte, Columbus, Dallas, Fort Worth-Arlington, Kansas City, Miami-Ft. Lauderdale, Milwaukee, Phoenix, Portland, Riverside-San Bernardino-Ontario, San Diego.

The 1998 AHS Metro Survey covered 15 metropolitan areas: Baltimore, Birmingham, Boston, Cincinnati, Houston, Minneapolis, Norfolk/Newport News, Oakland, Providence, Rochester, Salt Lake City, San Francisco, San Jose, Tampa, Washington, DC.

Date or Frequency:

The AHS Metro Survey was conducted annually between 1974 and 1996, and has been conducted biannually since 1998 with the exception of 2000 (a census year).

Aggregation:

No specific guidance regarding aggregation of the data exists in the extensive documentation on the AHS Metro Survey. The numbers of AI/AN/NA represented in each survey are sufficiently large to permit analyses for the specific metropolitan areas sampled in that year's data collection effort, so aggregation to compensate for small Ns is not necessary. If, however, the analyst wished to aggregate across data collection years to expand the number of MSAs in the data set, it would be advisable to use the AHS National data set (also, included in this catalog) rather than the Metro data set.

American Housing Survey: Metropolitan Surveys
(continued)

Data Collection Methodology: In-person interviews and telephone interviews are conducted by field interviewers.

Participation: Optional, without incentives

Response Rate: For 2004, the overall weighted response rate was 91 percent.

Sampling Methodology: To draw the AHS Metro Survey sample, the Census Bureau initially grouped the housing units enumerated in the 1990 Census of Population and Housing into census blocks and assigned these blocks to either the unit/group quarters frame or the area frame. Blocks located in an area that issued permits for new construction were assigned to the unit/group quarters frame. All other blocks were assigned to the area frame.

The unit/group quarters frame was then split into the unit frame and the group quarters frame by removing all group quarters and placing them in a separate frame. All housing units that were built after the 1990 census in areas where construction of new homes was monitored by building permits were placed into a separate frame, called the permit frame.

Sampling operations for all frames were performed separately within a designated group of counties in each state. Prior to the AHS Metro Survey sample selection, records selected by other Census Bureau surveys were removed from each of the frames to avoid having the same housing unit in sample for more than one survey. The Census Bureau selected the AHS Metro Survey sample from the remaining records.

Authorization: The U.S. Census Bureau conducts the American Housing Survey (AHS) to obtain up-to-date housing statistics for the Department of Housing and Urban Development (HUD). Title 12, Sections 1701Z-1 and 1701Z-2g of the U.S. Code authorize the Secretary of HUD to collect data from public and private agencies and protect the confidentiality of the data. Title 12, Section 1701Z-10 mandates the collection of the data for the AHS. This mandate covers the collection of data for the 2004 AHS Metro Survey.

Strengths: Data are collected on key policy issues, including housing. There are multiple years of data available. There are a large number of AI/ANs in the sample.

Limitations: Each data collection round of the AHS Metro Survey is limited to a relatively small number of metropolitan areas across the nation.

Access Requirements and Use Restrictions: Data are available to the public at no cost.

American Housing Survey: Metropolitan Surveys
(continued)

Contact Information:

The American Housing Survey Branch can be contacted by email at ahsn@census.gov or by phone at (301) 763-3235.

Older AHS Metro Survey data sets can be requested via HUDUSER (see http://www.huduser.org/datasets/ahs/ahsprev.html for list of specific datasets available).

Tables, reports and the actual data for 1998, 2002, and 2004 can be downloaded from: http://www.census.gov/hhes/www/housing/ahs/nationaldata.html.

Annual Survey of Jails (ASJ)

Sponsor: U.S. Department of Justice (DoJ)/Bureau of Justice Statistics (BJS)

Description: The Annual Survey of Jails (ASJ) provides an annual source of data on local jails and jail inmates. Data on the size of the jail population and selected inmate characteristics are obtained every five to six years from the Census of Jails (also profiled in this catalog). In each of the years between the full censuses, a sample survey of jails, the ASJ, is conducted to estimate baseline characteristics of the nation's jails and inmates housed in these jails. Data are supplied on inmate characteristics, admissions and releases, growth in the number of jail facilities, changes in their capacities and level of occupancy, growth in the population supervised in the community, changes in methods of community supervision, and crowding issues in local jails.

Relevant Policy Issues: Rates of Involvement with Justice System.

Data Type(s): Survey

Unit of Analysis: Jails

Identification of AI/AN/NA: On June 28, 2002, how many persons CONFINED in your jail facilities were:

- White, not of Hispanic origin
- Black or African American, not of Hispanic origin
- American Indian/Alaska Native (AI/AN)
- Asian
- Hispanic or Latino
- Native Hawaiian or other Pacific Islander (NH/PI)
- Other categories in your information system (Specify)

AI/AN/NA Population in Data Set: The nationally-representative sample included all public and private jails in selected jail jurisdictions and 50 regional jails. Each of these facilities reports on the number of AI/AN and NH/PI confined in their facility at midyear.

Geographic Scope: The geographic scope of the data is national. All 50 states and the District of Columbia are included. Analysis may be possible by state, county, and city depending on sample size within each subgroup.

Date or Frequency: The survey has been conducted annually since 1982, except every 5th year when the National Jail Census is conducted.

Data Collection Methodology: Data are collected primarily by mail with a web-based reporting option and telephone follow-up for nonrespondents.

Participation: Optional, without incentives

Annual Survey of Jails (ASJ)
(continued)

Response Rate: After follow-up phone calls, 100 percent of the jails provided data on key variables such as number of confined persons, number of male and female inmates by adult and juvenile, number of inmates by race and Hispanic origin, average daily population (ADP), and total rated capacity of jails. Data were not imputed for any items.

Sampling Methodology: Using information from the 1999 Census of Jails, a sample of jail jurisdictions was selected for the 2002 survey. The sample included all jail facilities (948) in 878 jurisdictions. Large jails and regional jails were in the sample with certainty. The remaining jurisdictions were stratified into two groups: jurisdictions with at least one juvenile inmate and jurisdictions holding adults only. Using stratified probability sampling, 474 jurisdictions were then selected from 10 strata based on the average daily population in the 1999 census. The sample selection was designed to precisely estimate the average daily population and one-day inmate population (i.e., highest population in the preceding month) for the entire nation.

Analysis: Standard errors are included in the documentation for some estimates but not for AI/AN or NH/PI estimates. These are included in an "other" category.

Strengths: Data are collected on a key policy issue, involvement with the justice system. There are multiple years of data available. The 2002 Annual Survey of Jails had a very high response rate (100 percent of jails provided data on critical items). The documentation is very detailed and readily accessible through the Internet.

Limitations: This is a facility-level rather than an individual-level data set. Counts of AI/AN or NH/PI persons being confined are available for the facilities, by state, and nationally. This information cannot be associated with individual characteristics for additional analysis.

Other: Researchers also may be interested in the Annual Survey of Jails in Indian Country (also described in this catalog).

Access Requirements and Use Restrictions: Data set is available to the public at no cost.

Contact Information: 2002 data and documentation can be downloaded at:
http://webapp.icpsr.umich.edu/cocoon/NACJD-STUDY/04428.xml.

Data archive information:
National Archive of Criminal Justice Data
ICPSR
University of Michigan
Institute for Social Research
P.O. Box 1248
Ann Arbor, MI 48106-1248

(800) 999-0960
(313) 763-5011
nacjd@icpsr.umich.edu
http://webapp.icpsr.umich.edu/cocoon/NACJD-SERIES/00007.xml

Questions for the Bureau of Justice Statistics should be addressed to:
James Stephan
Statistician
Bureau of Justice Statistics
810 Seventh Street, NW
Washington, DC 20531
USA
(202) 616-3289
James.Stephan@usdoj.gov or
askbjs@usdoj.gov

Behavioral Risk Factor Surveillance System (BRFSS)

Sponsor: U.S. Department of Health and Human Services (DHHS)/Centers for Disease Control and Prevention (CDC)

Description: The objective of the Behavioral Risk Factor Surveillance System (BRFSS) is to collect uniform, state-specific data on preventive health practices and risk behaviors that are linked to chronic diseases, injuries, and preventable infectious diseases in the adult population. Factors assessed by the BRFSS include tobacco use, health care coverage, HIV/AIDS knowledge and prevention, and physical activity.

Relevant Policy Issues: Measurement of Health Status.

Data Type(s): Survey

Unit of Analysis: Individual

Identification of AI/AN/NA: Question: Which one or more of the following would you say is your race?

- White
- Black/African American
- Asian
- Native Hawaiian/Pacific Islander (NH/PI)
- American Indian or Alaska Native (AI/AN)
- Other

If more than one response is given, the following question is asked for clarification: Which of these groups best represent your race?

- White
- Black/African American
- Asian
- Native Hawaiian/Pacific Islander (NH/PI)
- American Indian or Alaska Native (AI/AN)
- Other

AI/AN/NA Population in Data Set: In 2005, out of 356,112 total records, 6,904 respondents reported AI/AN as the race that best described them and 1,503 reported NH/PI as the race that best described them.

Geographic Scope: The geographic scope of the study is national. Over time, the number of states participating in the survey has increased, so that by 1994, 50 states, the District of Columbia, Puerto Rico, Guam, and the Virgin Islands were participating in the BRFSS.

Behavioral Risk Factor Surveillance System (BRFSS)
(continued)

The following geographic indicators are available on the public use file for analysis: state, county, zip code, indicator of residence within or outside a metropolitan statistical area.

Date or Frequency: Data collection has been conducted yearly since 1984.

Data Collection Methodology: BRFSS field operations are managed by state health departments, which follow guidelines provided by the CDC. These health departments participate in developing the survey instrument and conduct the interviews either in-house or through use of contractors. The data are transmitted to the CDC's National Center for Chronic Disease Prevention and Health Promotion's Behavioral Surveillance Branch for editing, processing, weighting, and analysis.

In 2005, all 53 states and territories used computer-assisted telephone interviewing (CATI). The core portion of the questionnaire lasts an average of 10 minutes. Interview time for modules and state-added questions is dependent upon the number of questions asked, but generally extend the interview period by an additional 5 to 10 minutes.

Participation: Optional, without incentives

Response Rate: Across the 53 states and territories, the median overall response rate in 2004 was 41.2 percent (minimum was 22.0 percent and maximum was 63.4 percent). The overall response rate assumes that all likely households are households and that 98 percent of known or probable households contain an adult who uses the phone number.

Sampling Methodology: In a telephone survey such as the BRFSS, a sample record is one telephone number in the list of all telephone numbers selected for dialing. In order to meet the BRFSS standard for participating states' sample designs, sample records must be justifiable as a probability sample of all households with telephones in the state. All participating areas met this criterion in 2004. Fifty–one projects used a disproportionate stratified sample design. Puerto Rico, Guam and the Virgin Islands used a simple random sample design.

Analysis: Overall approximately 95 percent of U.S. households have telephones, but coverage ranges from 87 to 98 percent across states and also varies for subgroups. People living in the South, minorities, and those in lower socioeconomic groups typically have lower telephone coverage. No direct method of compensating for non-telephone coverage is employed by the BRFSS; however, weighting adjustments for differences in probability of selection and non-response may compensate to some degree for non-telephone coverage.

Strengths: Data are collected on a key policy issue, health. Multiple years of data are available. Large sample sizes are available for AI/ANs and NH/PIs.

Behavioral Risk Factor Surveillance System (BRFSS)
(continued)

Limitations: Many states included in the study have significant AI/AN/NA populations that may not be reached through phone interviews because they do not have telephones.

Other: An edited and weighted data file is provided to each participating health department for each year of data collection, and summary reports of state-specific data are prepared by CDC. Health departments use the data for a variety of purposes, including identifying demographic variations in health related behaviors, targeting services, addressing emergent and critical health issues, proposing legislation for health initiatives, and measuring progress toward state and national health objectives.

Access Requirements and Use Restrictions: Data are available to public at no cost.

Contact Information:
Suzianne Ellington Garner, MPA
Deputy Chief, Behavioral Surveillance Branch
Division of Adult and Community Health
CDC
4770 Buford Hwy, MS K-66
Atlanta, GA 30341
(770) 488-6005
suzianne.garner@cdc.hhs.gov

Data can be accessed at http://www.cdc.gov/brfss.

Consumer Assessment of Healthcare Providers and Systems (CAHPS)
Health Plan Survey Response Data

Sponsor: U.S. Department of Health and Human Services (DHHS)/Agency for Healthcare Research and Quality (AHRQ)

Description: The Consumer Assessment of Healthcare Providers and Systems (CAHPS) program develops and supports the use of a family of standardized surveys that ask consumers and patients to report on and evaluate their experiences with health care. CAHPS surveys include ratings of personal doctors and other health care staff, as well as an overall rating of the health plan and ask patients and consumers to report on their experiences with health care services. CAHPS sponsors include various public and private organizations that fund the administration of a CAHPS survey by collecting data from consumers and patients of a particular health plan.

There are two data sets available to researchers. The core database consists of responses to the CAHPS Health Plan Survey. Participating sponsors submit these data at the individual respondent level in accordance with specifications provided by the CAHPS Database. All health plan identifiers are removed from the public use database. Respondent records contain a unique health plan ID so responses from each health plan can be grouped together, but health plans can not be identified. The data also do not include respondent names, addresses, telephone numbers, and member ID numbers. Certain survey administration data (e.g., mode of administration, survey language) and descriptive information (e.g., state) may also be included in this data set. The second database is the Survey Administration and Health Plan Characteristics Data. This data set includes information regarding survey administration, such as mode of administration, response rates, and dates of survey completion, as well as descriptive information relating to each of the sampled units (e.g., health plan products), such as type of organization, size of enrollment, tax status and ownership, and location.

Relevant Policy Issues: Measurement of Health Status and Factors Contributing to Measured Health Disparities.

Data Type(s): Survey

Unit of Analysis: Individual

Identification of AI/AN/NA: Race is self-reported, using the following question:

What is your race? Please mark one or more.

- White
- Black or African-American
- Asian
- Native Hawaiian or other Pacific Islander (NH/PI)
- American Indian or Alaska Native (AI/AN)
- Other

Consumer Assessment of Healthcare Providers and Systems (CAHPS)
Health Plan Survey Response Data
(continued)

AI/AN/NA Population in Data Set:

Number of Records per Year by Type of Plan

Adult Commercial Plans:
2005: Out of 123,272 records, 1,745 are AI/AN and 1,006 are NH/PI
2004: Out of 111,680 records, 1,745 are AI/AN and 942 are NH/PI
2003: Out of 114,063 records, 1,390 are AI/AN and 888 are NH/PI
2002: Out of 94,546 records, 1,286 are AI/AN and 744 are NH/PI
2001: Out of 165,500 records, 2,732 are AI/AN and 1,294 are NH/PI

Child Commercial Plans:
2005: Out of 2,661 records, 28 are AI/AN and 9 are NH/PI
2004: Out of 7,024 records, 77 are AI/AN and 29 are NH/PI
2003: Out of 1,866 records, 30 are AI/AN and 2 are NH/PI
2002: Out of 5,600 records, 106 are AI/AN and 22 are NH/PI
2001: Out of 9,913 records, 153 are AI/AN and 118 are NH/PI

Adult Medicaid Plans:
2005: Out of 32,115 records, 1,011 are AI/AN and 194 are NH/PI
2004: Out of 59,515 records, 1,982 are AI/AN and 1,188 are NH/PI
2003: Out of 39,275 records, 1,143 are AI/AN and 211 are NH/PI
2002: Out of 48,109 records, 1,942 are AI/AN and 1,681 are NH/PI
2001: Out of 45,127 records, 1,596 are AI/AN and 1,875 are NH/PI

Child Medicaid Plans:
2005: Out of 40,204 records, 1,184 are AI/AN and 1,354 are NH/PI
2004: Out of 86,159 records, 2,291 are AI/AN and 806 are NH/PI
2003: Out of 31,082 records, 1,153 are AI/AN and 1,079 are NH/PI
2002: Out of 60,534 records, 2,018 are AI/AN and 472 are NH/PI
2001: Out of 36,940 records, 1,322 are AI/AN and 268 are NH/PI

State Children's Health Insurance Program (SCHIP):
2005: Out of 1,252 records, 5 are AI/AN and 4 are NH/PI
2004: Out of 16,657 records, 359 are AI/AN and 143 are NH/PI
2003: Out of 19,061 records, 402 are AI/AN and 132 are NH/PI
2002: Out of 18,910 records, 349 are AI/AN and 203 are NH/PI
(SCHIP data were not collected in 2001.)

Geographic Scope:

The geographic scope of CAHPS is national. The data files also include a state indicator. Some state-level analysis may be possible, but will depend on the number of records available for each state. State coverage varies from year to year, depending on which providers submit data for inclusion in the CAHPS Database. Details of the number of records per state are available in the CAHPS Health Plan Survey Chartbook, which is released on an annual basis and can be accessed at:
https://www.cahps.ahrq.gov/content/NCBD/Chartbook/2005_Chartbook.pdf

Consumer Assessment of Healthcare Providers and Systems (CAHPS)
Health Plan Survey Response Data
(continued)

Date or Frequency: Schedules for collecting CAHPS data vary by sponsor. The CAHPS Database is compiled every year, collecting data from health plans that have completed data collection in the past 12 months. Data are currently available from 2001 to 2005 (SCHIP data are available from 2002 to 2005). Data from 1998 to 2000 also exist in the CAHPS Database, but the quality of the data cannot be assured, and the CAHPS staff advises against using these data.

Beginning in 2006, the CAHPS Database will be expanded to include CAHPS Hospital Survey data, and will eventually be further expanded to include CAHPS Clinician and Group Survey Data. These survey data will all include unique respondent identifiers as well.

Aggregation: Researchers who are interested in combining multiple years of data for aggregation should consider that the basic reporting unit for CAHPS is the provider, even though the survey is at the individual respondent level. Across multiple years there will be differences in the providers submitting data; for example, in one year more large providers may submit and in another more Western providers. An aggregated data set will not necessarily be representative of the population by year or in combination. These differences in providers by year cannot be described, as the provider-level unique ID number is not consistently assigned from year to year. Also, each year, the data from the CAHPS surveys are case-mix adjusted in order to create the CAHPS benchmark. Researchers should consider re-running the case-mix adjustment (called the CAHPS Macro) on data aggregated across multiple years.

Data Collection Methodology: Data collection methodology varies by the CAHPS sponsor or vendor administering the CAHPS survey. Some information on the mode of data collection is included in the public use data set (e.g., each completed survey is coded either M = mail complete, T = telephone complete, or I = Internet complete).

Additionally, the CAHPS Program provides the following guidelines for sponsors concerning data collection: "A mixed-mode data collection protocol involving both mail and telephone is more likely to achieve a desired response rate than will either mode alone. Research conducted by the CAHPS grantees shows that differences in the types of responses collected by these different modes are minimal (these differences are called “mode effects”), so telephone and mail can be used together with confidence."

Participation: Optional, without incentives

Response Rate: Response rates are calculated and provided by the different CAHPS survey vendors and sponsors who submit data to the CAHPS Database. The 2005 CAHPS Database self-reported response rates vary by sponsor and range from 14 percent to 71 percent. The breakdown of response rates by population type are as follows:

Consumer Assessment of Healthcare Providers and Systems (CAHPS)
Health Plan Survey Response Data
(continued)

Adult Commercial: 19% - 71%
Adult Medicaid: 17% - 59%
Child Medicaid: 14% - 50%

Sampling Methodology: Sampling methods for CAHPS vary by sponsor. CAHPS provides guidelines for selecting a sample, including determining eligibility, calculating the estimated sample size needed for general reporting, and creating a frame of all covered lives or sampling policyholders only.

Strengths: Some CAHPS data sets contain a large number of AI/AN/NA respondents. Data are collected on key policy issues, including self-reported health status. There are multiple years of data available.

Limitations: The CAHPS survey is not administered in a consistent fashion. Instead, the CAHPS Database is a collection of surveys administered at the level of health plans. As such, not all health plans participate each year, so the mix of plans will vary across years. Additionally, sampling and data collection methods vary by plan, as plans hire vendors to administer the survey and these methods will vary by vendor.

Access Requirements and Use Restrictions: To access these data, researchers must submit a data release agreement and a description of the proposed research, as well as IRB clearance documentation.

Contact Information: Contact Dale Shaller, Managing Director, with any questions about the CAHPS Database and data requests.
Email: d.shaller@comcast.net
Phone: (651) 430-0759

Send proposals for data access and signed Data Use Agreements to:
CAHPS Project Staff
Westat
1650 Research Boulevard
RA1159
Rockville, MD 20850
Fax: (301) 251-1500

The CAHPS Database website is located at:
https://www.cahps.ahrq.gov/content/ncbd/ncbd_Intro.asp?p=105&s=5

Reports of Interest: Research Brief: *Racial and Ethnic Disparities in the Experiences of Health Care Consumers.* Published in November 2005 by the National CAHPS Benchmarking Database, under AHRQ Contract Number 290-0I-0003. Written by Karen Onstad.

California Health Interview Survey (CHIS)

Sponsor: UCLA Center for Health Policy Research

Description: The California Health Interview Survey (CHIS) is a biennial telephone survey that began in 2001. CHIS collects information on California children (0-11 yrs), adolescents (12-17 yrs), and adults (18 and older) about their health and health care access. Specific topics addressed by the survey include health status, health conditions, health-related behaviors, health insurance coverage, access to and use of health care services, and the health and development of children and adolescents. CHIS data are used to produce population-based estimates for most California counties, all major ethnic groups, and several ethnic subgroups within California. The overall sample is representative of the state's non-institutionalized population.

Relevant Policy Issues: Measurement of Health Status, Disease-specific Measurements, Key Health Disparities, Factors Contributing to Measured Health Disparities, Measures of Well-being for Children, and Measures of Well-being for Elders.

Data Type(s): Survey

Unit of Analysis: Individual

Identification of AI/AN/NA: The CHIS interview asks several questions about race. Below are those relevant to the identification of AI/AN/NA in this data source.

1. Would you describe yourself as Native Hawaiian, Other Pacific Islander, American Indian, Alaska Native, Asian, Black, African American, or White? (Interviewers are instructed to code all that apply.)

- White
- Black or African American
- Asian
- American Indian or Alaska Native (AI/AN)
- Other Pacific Islander (OPI)
- Native Hawaiian (NH)
- Other (Specify)

2. You said, American Indian or Alaska Native, and what is your tribal heritage? If you have more than one tribe, tell me all of them.

3. Are you an enrolled member in a federally or state recognized tribe?

4. Which tribe are you enrolled in?

5. You said you are Pacific Islander. What specific ethnic group are you, such as Samoan, Tongan, or Guamanian? If you are more than one, tell me all of them.

California Health Interview Survey (CHIS)
(continued)

AI/AN/NA Population in Data Set:

The CHIS 2003 Random Digit Dial (RDD) sample is representative of California's non-institutionalized population. The numbers below are based on any mention of a specific race or ethnicity (rather than the single-race categories):

Adults
Total number of records: 42,044
AI/AN, Hispanic: 740
AI/AN, Non-Hispanic: 1,157
NH/PI, Hispanic: 61
NH/PI, Non-Hispanic: 199

Adolescents (ages 12-17)
Total number of records: 4,010
AI/AN, Hispanic: 212
AI/AN, Non-Hispanic: 153
NH/PI, Hispanic: 27
NH/PI, Non-Hispanic: 39

Children (ages 0-11)
Total number of records: 8,156
AI/AN, Hispanic: 195
AI/AN, Non-Hispanic: 175
NH/PI, Hispanic: 34
NH/PI, Non-Hispanic: 57

AI/AN/NA Subpopulations:

Data source allows identification of federally recognized tribes: Apache, Blackfeet, Cherokee, Choctaw, Mexican American Indian, Navajo, Pomo, Pueblo, Sioux, Yaqui.

Additional subpopulation categories include Native Hawaiian alone, Pacific Islander alone, and a breakdown by Pacific Island (i.e., Samoan/American, Samoan, Guamanian, Tongan, Fijian, Polynesian, and Other Pacific Islander). Sample size for some Tribes or Islands may be too small for analysis.

Geographic Scope:

The geographic scope of the study is the state of California. Additional geographic identifiers include counties, zip codes, and exact longitudes and latitudes of the residence for about 80 percent of the sample.

Date or Frequency:

The CHIS survey began in 2001 and is fielded every 2 years. Data from 2003 are the most recent data that are publicly available.

Data Collection Methodology:

Data are collected using a computer assisted telephone interviewing (CATI) system, with respondents selected through a geographically stratified random digit dialing approach. One adult in a household is selected and responds for him/herself and one sampled child (0-11 yrs) if there are any in the household. Adolescents respond for themselves, after approval is given from their guardian. Up to three individuals in any given household may be sampled (adult,

adolescent, child). In addition to English, the survey is fielded in five additional languages (Spanish, Mandarin, Cantonese, Korean and Vietnamese).

Participation: Optional, without incentives

Response Rate: Most recent response rate was 33.5 percent (composite of screener and interview completion rates).

Sampling Methodology: CHIS employs a multi-stage sample design in which the state of California is divided into 41 geographic sampling strata (primarily counties) and within each geographic stratum, households were selected through random-digit dialing. Within each household, an adult (age 18 and over) respondent was randomly selected, and for those households with adolescents (ages 12-17) and children (under age 12), one of each was randomly selected for interview.

Analysis: The complex survey design of the CHIS requires that adjustments to weighting and standard error calculations be made in order to produce robust estimates. Failure to make these adjustments can yield estimates where the standard error appears smaller than it actually should be, suggesting the accuracy of the estimate is better than it actually is. The technique used to address this adjustment process in the CHIS Public Use Files is replication, and the data set includes the set of replicate weights for users to apply in the calculation of standard errors. Special software is required to conduct statistical analyses using replicate weights.

Strengths: The CHIS data sets contain a large number of AI/AN/NA respondents. The data are collected on key policy issues, including health and child welfare. There are multiple years of data available. The sample size of the overall survey is of sufficient size to allow for analyses of racial/ethnic subgroups of interest, which includes a breakout of Native Hawaiian from Pacific Islander as well as specific tribal affiliations. While sample sizes for certain tribes are small, some tribes do have sufficient sample size to permit at least basic descriptive statistical analyses. Furthermore, many of the questions fielded in the CHIS are taken from the National Health Interview Survey, which can provide a national benchmark for many variables.

Limitations: While a very large survey, results are only generalizable to the state of California. The sample does not include an institutionalized population or people without telephones. Some tribes and other subpopulations have very small sample size that will allow only minimal analyses. The response rate is low.

Access Requirements and Use Restrictions: CHIS Public Use Files are available through a data use agreement at no cost.

California Health Interview Survey (CHIS)
(continued)

Contact Information:

The main website for the survey is located at http://www.chis.ucla.edu/. Data can be downloaded at this site. There is also an on-line analysis query system named "AskCHIS." In addition, potential users can link to a wide variety of other CHIS-related information at this site.

California Health Interview Survey
UCLA Center for Health Policy Research
10960 Wilshire Blvd, Suite 1550
Los Angeles, CA 90024
Toll free: (866) 275-2447
E-mail: chis@ucla.edu

Census 2000

Sponsor: U.S. Department of Commerce/U.S. Census Bureau

Description: The Decennial Census occurs every 10 years to count the population and housing units for the entire United States. Its primary purpose is to provide the population counts that determine how seats in the U.S. House of Representatives are apportioned. The U.S. Census Bureau provides three types of data products that may be useful to the interested researcher:

- Tabular data in the form of summary files,
- Raw data in the form of Public Use Microdata Sample files (PUMS files), and
- Census briefs and special reports.

Tabular data: Summary files

The U.S. Census Bureau has released a series of summary files that present Census 2000 data in tabular form. The primary summary files are:

1. Summary File 1: This file contains 286 detailed tables focusing on age, sex, households, families, and housing units. These tables provide in-depth figures by race and Hispanic origin. Counts are also provided for over 40 American Indian and Alaska Native tribes and for groups within race categories. The race categories also include 12 Native Hawaiian and other Pacific Islander groups.

2. Summary File 2: This file contains 47 detailed tables focusing on age, sex, households, families, and occupied housing units for the total population. These tables are repeated for 249 detailed population groups, including American Indian, Alaskan Native, 9 Native Hawaiian or other Pacific Islander groups, and 40 American Indian and Alaska Native tribes. For each of these groups, data are provided for that group alone and in combination with one or more other races.

3. Summary File 3: This file consists of 813 detailed tables of Census 2000 social, economic and housing characteristics compiled from a sample of approximately 19 million housing units (about 1 in 6 households) that received the Census 2000 long-form questionnaire. Fifty-one tables are repeated for 9 major race categories including American Indian/Alaska Native and Native Hawaiian/Other Pacific Islander.

4. Summary File 4: This file consists of 213 population tables and 110 housing tables. Each table is repeated for 336 population groups: the total population, 132 race groups, 78 American Indian and Alaska Native tribe categories (reflecting 39 individual tribes), and 9 Native Hawaiian or other Pacific Islander groups.

5. The American Indian and Alaska Native Summary File (AIANSF): These sample data are presented in 213 population tables and 110 housing tables. The tables are repeated for the total population, the total

American Indian and Alaska Native population, the total American Indian population, the total Alaska Native population, and for 1,081 additional self-reported American Indian and Alaska Native tribes and villages without consideration of any designation of federal or state recognition. **(Please note that the AIANSF is profiled separately in this data catalog.)**

Raw data: 1 percent and 5 percent PUMS files
The PUMS files contain records of households, people, or housing units with identifying information removed and other precautions taken to prevent the violation of confidentiality. PUMS files often show data only for identified geographic areas (such as states) that meet a certain population threshold. The 1 percent PUMS files have state-level Census 2000 data containing individual records of the characteristics for a 1 percent sample of people and housing units. The 5 percent PUMS files contain similar information for a 5 percent sample of people and housing units.

Census briefs and special reports
The Census 2000 Brief series focus on discussing key topics covered by the Census and exploring the geographic distribution of the topics. The Census 2000 Special Report series provides an in-depth analysis of Census 2000 population and housing topics.
Examples of census briefs and special reports of particular interest include:

- *The American Indian and Alaska Native Population: 2000* (C2KBR/01-15)
- *American Indian and Alaska Native Tribes for the United States, Regions, Divisions, and States* (PHC-T-18)
- *We the People: American Indians and Alaska Natives in the United States* (CENSR-28)
- *The Native Hawaiian and other Pacific Islander Population: 2000* (C2KBR/01-14)

Relevant Policy Issues: Demographic and Economic Indicators, Measurement of Health Status, Income Status, Unemployment Rates, Economic Assistance Program Participation Rates, Educational Attainment, Measures of Well-being for Families/Households, Factors Contributing to Well-being Disparities of Families, Housing Quality, Type of Housing, Housing Ownership, Rental Unit Quality and Cost, and Transportation Availability.

Data Type(s): Census survey

Unit of Analysis: Individual

Identification of AI/AN/NA: As described above, identification of the AI/AN/NA population differs across the many Census 2000 data products. Some data products present the AI/AN/NA population into 2 groups: American Indian/Alaskan Native and Native Hawaiian/Other Pacific Islander while other data products provide more detailed

breakdowns (e.g., distinction between American Indians and Alaska Native, distinction between Native Hawaiian and other Pacific Islanders). Some data products provide detailed breakdowns of these groups presenting data for different tribal affiliations and different Pacific Islander groups. For example, the PUMS files, the AIANSF file, and Summary Files 2 and 4 present tribal affiliation data.

Please note that the tribal affiliation data reflect the written entries by respondents, who identified themselves as AI/AN, and provided an entry for their enrolled or principal tribe or village. Some of the responses (for example, Colorado River and Village of Alakanuk) represent reservations or native villages. The information on tribe or village is based on self-identification without consideration of any designation of federal or state recognition.

AI/AN/NA Population in Data Set:

The following counts are reported in *Profiles of General Demographic Characteristics: 2000 Census of Population and Housing*. These data are based on 100 percent counts derived from the Census 2000 short form:

Total population: 281,421,906
American Indian and Alaska Native: 2,475,956
Native Hawaiian: 140,652
Guamanian or Chamorro: 58,240
Samoan: 91,029
Other Pacific Islander: 108,914

AI/AN/NA Subpopulations:

As described in detail above, some data products provide detailed breakdowns (e.g., distinction between American Indians and Alaska Native, distinction between Native Hawaiian and other Pacific Islanders) and some present data for different tribal affiliations and different Pacific Islander groups.

Geographic Scope:

The geographic scope of the study is national. Geographic areas covered by the data in the Census 2000 summary files include:

- Region (e.g., Midwest, Northeast, South, West)
- Division (e.g., East North Central, East South Central, Middle Atlantic, Mountain, New England)
- State
- County (county subdivision, census tract)
- Place (cities, towns, municipalities)
- Consolidated cities
- American Indian Area/Alaska Native Area/Hawaiian Homeland (including reservations or statistical entities, off-reservation trust lands, Hawaiian homelands, tribal census tracts, tribal subdivisions and remainders)
- Alaska Native Regional Corporation (e.g., Ahtna Alaska Native Regional Corporation, Aleut Alaska Native Regional Corporation, Arctic Slope Alaska Native Regional Corporation)
- Metropolitan Statistical Area, Primary Metropolitan Statistical Area
- New England County Metropolitan Area

- Urban areas

Geographic areas covered by the Census 2000 PUMS data include:

- Region (e.g., Midwest, Northeast, South, West)
- Division (e.g., East North Central, East South Central, Middle Atlantic, Mountain, New England)
- State
- Public Use Microdata Area Code (PUMA)
- Super Public Use Microdata Area Code (SuperPUMA)
- Metropolitan Area (MA): MSA/CMSA for PUMA and SuperPUMA

Date or Frequency: The Census is conducted every 10 years in years ending in zero. The next census is scheduled for 2010. Data are available for each year since 1790. American Indians were first enumerated as a separate group in the 1860 Census. The 1890 census was the first to count American Indians, including some tribes, throughout the country.

Data Collection Methodology: Census 2000 data were collected by mail, telephone, personal interview, and Internet.

Participation: Mandatory

Response Rate: The national final response rate for Census 2000 was 67 percent and represents responses received by mail, telephone or over the Internet through September 7, 2000. The final response rates for 117 American Indian Areas are listed at the following website: http://www.census.gov/dmd/www/response/disp-fro-res.txt.

Sampling Methodology: Basic demographic and housing questions (for example, race, age, and relationship to householder) were asked for every person in all housing units in the United States. A sample of housing units was also selected to receive more detailed questions in the long form of Census 2000, containing items such as income, occupation, and housing costs. The sampling unit for the long form Census 2000 was the housing unit, including all occupants. There were four different housing unit sampling rates: 1-in-8, 1-in-6, 1-in-4, and 1-in-2 (designed for an overall average of about 1-in-6). The Census Bureau assigned these varying rates based on precensus occupied housing unit estimates of various geographic and statistical entities, such as incorporated places and interim census tracts. For people living in group quarters or those enumerated at long-form-eligible service sites (shelters and soup kitchens), the sampling unit was the person and the sampling rate was 1-in-6.

Analysis: Detailed information regarding the design effects and standard errors for each of the 2000 Census PUMS files and the summary files is available for download from online links on the Census 2000 website.

Authorization: The Census is mandated by the U.S. Constitution, Article I, Section 2. Participation in the Census is required by law set forth in Sections 141 and 193 of Title 13 of the United States Code.

Census 2000
(continued)

Strengths:	The Census 2000 PUMS files and summary files contain a representative sample of the AI/AN population and selected tribes and villages. The Census 2000 PUMS files and summary files also contain a representative sample of the NH/PI population with some data products providing information on select OPI groups. Data are collected on key policy issues. The U.S. Census Bureau's website provides extremely comprehensive documentation on the methodology, results, and interpretation of census data.
Limitations:	The Census 2000 PUMS files are a very large set of complex files. Considerable expertise in working with data of these types will likely be required. The summary files are less complex but more numerous, thus finding the particular table(s) of interest may be challenging.
Access Requirements and Use Restrictions:	Both the PUMS data and the summary files are available to the public at no cost.
Contact Information:	A U.S. Census Bureau list of contacts by subject area is available at the following website: http://www.census.gov/contacts/www/c-census2000.html. The Census 2000 summary files as well as supporting documentation are available at the U.S. 2000 Census website: http://factfinder.census.gov/servlet/DatasetMainPageServlet?_program=DEC&_submenuId=datasets_1&_lang=en.

Census 2000 - The American Indian and Alaska Native Summary File

Sponsor: U.S. Department of Commerce/U.S. Census Bureau

Description: As mandated by the U.S. Constitution, the Decennial Census is conducted every 10 years to count the population and housing units for the entire United States.

The American Indian and Alaska Native Summary File (AIANSF) is based on Census 2000 data. Data from Census 2000 on the American Indian and Alaska Native population (AI/AN) are derived from a limited number of basic questions asked of the entire American Indian and Alaska Native population and every corresponding housing unit (referred to as the 100-percent questions found on the "short form"), and from additional questions asked of a sample of the population and housing units (referred to as the sample questions, found on the "long form"). The AIANSF provides sample data based on both the 100-percent and the sample questions.

Data in the AIANSF include, for example, age, Hispanic or Latino origin, household relationship, sex, educational attainment, veteran status, income and poverty status, housing tenure (owner-or renter-occupied), physical housing characteristics, and mortgage and rental cost characteristics. These data are available for the total AI/AN population, the total American Indian population, the total Alaska Native population, and for 1,081 self-reported AI/AN tribes or villages without consideration of any designation of federal or state recognition.

Relevant Policy Issues: Demographic and Economic Indicators, Measurement of Health Status, Income Status, Unemployment Rates, Economic Assistance Program Participation Rates, Educational Attainment, Measures of Well-being for Families/Households, Factors Contributing to Well-being Disparities of Families, Housing Quality, Type of Housing, Housing Ownership, Rental Unit Quality and Cost, Transportation Availability.

Data Type(s): Census survey

Unit of Analysis: Individual

Identification of AI/AN/NA: Tribal data provided in the AIANSF reflect the written entries by respondents, who identified themselves as AI/AN, and provided an entry for their enrolled or principal tribe or village. Some of the responses (for example, Colorado River and Village of Alakanuk) represent reservations or native villages. The information on tribe or village is based on self-identification without consideration of any designation of federal or state recognition.

The listing of American Indian and Alaska Native tribes and villages is derived from the American Indian Tribal Detailed Classification List for the 1990 census, which was expanded to include individual Alaska Native villages, when provided as a written response to the question on race. The list was updated based on a December 1997 Federal Register Notice, entitled *Indian Entities Recognized and*

Eligible to Receive Service From the United States Department of Interior, Bureau of Indian Affairs, issued by the Office of Management and Budget. The list of 1,081 tribes or villages for which summary tables are available can be found in the technical report for the AIANSF sample file available at: http://www.census.gov/prod/cen2000/doc/aiansf.pdf.

AI/AN/NA Population in Data Set: Unweighted counts of AI/AN/NA in the AIANSF are not available, but given the scope of the Census, the counts are expected to be sufficiently high to support most analyses.

AI/AN/NA Subpopulations: The AIANSF allows identification of members of federally and state-recognized tribes and villages by tribe or village. Subpopulations are identified by specific affiliation.

Geographic Scope: The geographic scope of the study is national. Geographic areas covered by the data in the AIANSF include:

- Region (e.g., Midwest, Northeast, South, West);
- Division (e.g., East North Central, East South Central, Middle Atlantic, Mountain, New England);
- State;
- American Indian Area/Alaska Native Area/Hawaiian Homeland (e.g., Acoma Pueblo and Off-Reservation Trust Land, and Agua Caliente Reservation);
- Alaska Native Regional Corporation (e.g., Ahtna Alaska Native Regional Corporation, Aleut Alaska Native Regional Corporation);
- Metropolitan Statistical Area/Consolidated Metropolitan Statistical Area;
- Primary Metropolitan Statistical Area; and
- New England County Metropolitan Area.

Date or Frequency: The Census is conducted every ten years in years ending in zero. The next census is scheduled for 2010. Data are available for each year since 1790. American Indians were first enumerated as a separate group in the 1860 Census. The 1890 census was the first to count American Indians, including some tribes, throughout the country.

Data Collection Methodology: Census 2000 data were collected by mail, telephone, personal interview, and Internet.

Participation: Mandatory

Response Rate: The national final response rate for Census 2000 was 67 percent and represents responses received by mail, telephone or over the Internet through September 7, 2000. The final response rates for 117 American Indian Areas are listed at the following website: http://www.census.gov/dmd/www/response/disp-fro-res.txt.

Census 2000 - The American Indian and Alaska Native Summary File (continued)

Sampling Methodology: Every person and housing unit in the United States was asked basic demographic and housing questions (for example, race, age, and relationship to householder). A sample of these people and housing units was asked more detailed questions about items such as income, occupation, and housing costs. The sampling unit for Census 2000 was the housing unit, including all occupants. There were four different housing unit sampling rates: 1-in-8, 1-in-6, 1-in-4, and 1-in-2 (designed for an overall average of about 1-in-6). The Census Bureau assigned these varying rates based on precensus occupied housing unit estimates of various geographic and statistical entities, such as incorporated places and interim census tracts. For people living in group quarters or those enumerated at long-form-eligible service sites (shelters and soup kitchens), the sampling unit was the person and the sampling rate was 1-in-6.

Analysis: Detailed information regarding the design effects and standard errors for the 2000 Census long form is available from the following publication: *Summary File 4- 2000 Census of Population and Housing: Technical Documentation (Chapter 8)* and *American Indian and Alaska Native Summary File Technical Documentation (Chapter 8).*

Authorization: The Census is mandated by the U.S. Constitution, Article I, Section 2. Participation in the Census is required by law set forth in Sections 141 and 193 of Title 13 of the United States Code.

Strengths: The AIANSF data set contains a representative sample of the AI/AN population and selected tribes and villages. Data are collected on key policy issues. The U.S. Census Bureau's website provides extremely comprehensive documentation on the methodology, results, and interpretation of census data.

Limitations: The AIANSF is a very large set of complex files. Considerable expertise in working with data of this type will likely be required.

Access Requirements and Use Restrictions: Data are available to the public at no cost.

Contact Information: No specific contact information regarding the AIANSF is given, however a U.S. Census Bureau list of contacts by subject area is available at the following website: http://www.census.gov/contacts/www/c-census2000.html.

The AIANSF data as well as supporting documentation are available at the U.S. 2000 Census website: http://factfinder.census.gov/servlet/DatasetMainPageServlet?_program=DEC&_submenuId=datasets_1&_lang=en.

Census of Agriculture (2002)

Sponsor: U.S. Department of Agriculture/National Agriculture Statistics Service

Description: The Census of Agriculture provides periodic and comprehensive statistics about agricultural operations, production, operators and land use for 1992, 1997, and 2002. Agricultural statistics are used by government, businesses, and other institutions. Federal, state, and local agencies use data for planning rural development, extension work, and agricultural research. The census is the only source of detailed, complete, consistent agricultural data for each county; it also includes such data for the states and the United States.

Relevant Policy Issues: Economic Opportunity and Measurement of Economic/Employment Disparities between AI/AN/NA and General Population.

Data Type(s): Census survey

Unit of Analysis: Principal Operator, Farm, and Ranch

Identification of AI/AN/NA: The race categories collected for the Census of Agriculture are:

- White
- Black or African American
- American Indian or Alaska Native - specify tribe (AI/AN)
- Native Hawaiian or other Pacific Islander (NH/PI)
- Asian

AI/AN/NA Population in Data Set: Total number of farm operators: Approximately 2,464,000
AI/AN principal farm operators: Approximately 5,268
NH/PI principal farm operators: Approximately 280

Note: These “unweighted” counts were calculated by taking the values reported in *Operators by Race, Special Reports Part 1, 2002 Census of Agriculture* and dividing them by the approximate weights for nonresponse adjustment and coverage adjustment (0.34 for AI/AN and 0.285 for NH/PI).

Although self-reported tribal affiliation is collected on the Census of Agriculture, the data set is not available to the public and it is not clear whether analyses employing that information could be made available via a "Special Tabulations" request to the National Agricultural Statistics Service.

Geographic Scope: The geographic scope of the Census of Agriculture is national. Geographic analysis also is possible by state and county.

Date or Frequency: The Census of Agriculture is planned for every 5 years. It was conducted in 1992, 1997, and 2002. Reports and tabulations are available for each of these data collection efforts.

Census of Agriculture (2002)
(continued)

Data Collection Methodology:	The Census of Agriculture is a mail survey with telephone and face-to-face interviewing follow-up for nonrespondents.
Participation:	Mandatory
Response Rate:	Multiple response rates were calculated, but these have not been published. Appendix A of the *2002 Census of Agriculture Volume 1 Chapter 1: U.S. National Level Data* report lists a minimal response rate of 75 percent.
Sampling Methodology:	Appendix C of the *2002 Census of Agriculture Volume 1 Chapter 1: U.S. National Level Data* report states that "all name and address records on the final [Census Mail List] received a 2002 Census of Agriculture report form."
Authorization:	Title 7, Chapter 55, 2204g. Authority of Secretary of Agriculture to conduct Census of Agriculture.
Strengths:	Strengths of the data source include sufficient numbers of members of the AI/AN/NA population. Moreover, special efforts were undertaken in the 2002 Census of Agriculture to address representation of AI/AN/NA farm operators. In addition, there are multiple years of data available.
Limitations:	The Census of Agriculture only provides indirect measures of economic well-being (e.g., measures of size of farms, productivity of farms, type of produce, livestock, etc. produced by the farm).
Other:	The U.S. Department of Agriculture also conducted a pilot project in conjunction with the 2002 Census of Agriculture to collect agricultural census data for farms and ranches on American Indian reservations in Montana, North Dakota, and South Dakota. This is the first time agricultural census data for American Indian reservations based on individual farm and ranch reports have ever been published by the National Agricultural Statistics Service (NASS) of the U.S. Department of Agriculture. The results of this pilot project have been published in *American Indian Reservations: Montana, North Dakota, and South Dakota. Pilot Project. Specialty Products, Part 1. AC-02-SP-1*. This report is available at http://www.nass.usda.gov/census/amindian.pdf. It is important to note that the methodology used to account for AI/AN/NA farm operators in this pilot project differs from that used in the overall Census of Agriculture. The pilot project emphasized individual-level reports while the overall Census uses both individual-level reports as well as aggregated information obtained from reservation-level reports.
Access Requirements and Use Restrictions:	Data set is not available to the public, but interested parties can request analyses. Also, published tables and reports are available. Special Tabulations are publishable, resummarized data tables from the Census of Agriculture or NASS surveys. Requests for Special Tabulations are considered

when the requested data are not published elsewhere. Depending on the complexity of the request, specialized analyses may be done for no or minimal cost. More complex requests are chargeable and the minimum charge is $500 for a Special Tabulation.

Contact Information:

Agriculture Statistics Hotline (800) 727-9540
National Agricultural Statistics Service
U.S. Department of Agriculture
USDA-NASS
1400 Independence Ave, SW
Washington, DC 20250

Census of Jails

Sponsor: U.S. Department of Justice (DoJ)/Bureau of Justice Statistics (BJS)

Description: The 1999 Census of Jails is the seventh in a series of data collection efforts aimed at studying the nation's locally administered jails. The 1999 census enumerated 3,365 locally administered confinement facilities that held inmates beyond arraignment and were staffed by municipal or county employees and 11 facilities maintained by the Federal Bureau of Prisons that functioned as jails. Variables include information on jail population by legal status (i.e., convicted or not convicted), age and sex of prisoners, maximum sentence, admissions and releases, available services and programs, structure and capacity, facility age and use of space, expenditure (i.e., per diem fees charged/paid for confining inmates), employment, staff information, and inmate health issues, which includes statistics on drugs, AIDS, and tuberculosis.

Relevant Policy Issues: Rates of Involvement with Justice System.

Data Type(s): Census survey

Unit of Analysis: Correctional facility

Identification of AI/AN/NA: On June 30, 1999, how many persons CONFINED in your jail facilities were:

- White, not of Hispanic origin
- Black or African American, not of Hispanic origin
- American Indian/Alaska Native (AI/AN)
- Asian
- Hispanic or Latino
- Native Hawaiian or other Pacific Islander (NH/PI)
- Other

On June 30, 1999, how many staff employed by your jail facility were:

- White, not of Hispanic origin
- Black or African American, not of Hispanic origin
- American Indian/Alaska Native (AI/AN)
- Asian
- Hispanic or Latino
- Native Hawaiian or other Pacific Islander (NH/PI)
- Other

-

Census of Jails
(continued)

Of all CORRECTIONAL OFFICERS reported in item 23b [staff count item], how many were:

- White, not of Hispanic origin
- Black or African American, not of Hispanic origin
- American Indian/Alaska Native (AI/AN)
- Asian
- Hispanic or Latino
- Native Hawaiian or other Pacific Islander (NH/PI)
- Other

AI/AN/NA Population in Data Set: The 1999 Census includes information on 3,084 jail jurisdictions. Each of these jurisdictions reports on the number of AI/AN and NH/PI confined in their facility(s) at mid-year.

Geographic Scope: The geographic scope is national. This is a census of jail facilities in 46 states and the District of Columbia. Connecticut, Delaware, Hawaii, Rhode Island, and Vermont are excluded because they operate combined jail-prison facilities. Analysis is possible by state, county, census area, zip code or groups of zip codes, and individual jail jurisdictions.

Date or Frequency: The Census of Jails, previously known as the National Jail Census, is conducted every 5 to 6 years for the U.S. Department of Justice, Bureau of Justice Statistics (BJS) by the U.S. Census Bureau. Censuses have been conducted in 1970, 1972, 1978, 1983, 1988, 1993 and 1999. Data collection for the next census was conducted in 2005 and 2006. These data should be available in 2007 or early 2008.

Data Collection Methodology: The mailing list used for the Census of Jails is derived from a facility list maintained by the Census Bureau for BJS, correctional association directories, and other secondary sources. Census forms were mailed to facilities. In addition to a paper form, BJS offered respondents in large jurisdictions an electronic version via the Internet, which allowed them to complete and submit their questionnaire on-line. Follow-up included additional mail and fax requests and repeated telephone contacts.

Participation: Optional, without incentives

Response Rate: Data were obtained by mailed and web-based survey questionnaires. After followup phone calls, nearly 100 percent of jails provided critical data items such as gender of inmates held and number of inmates on June 30, 1999.

Analysis: Because there was nonresponse and incomplete data from a small number of facilities on non-critical items, survey staff estimated totals for their reporting and imputed data for some missing non-critical items. Full documentation of these procedures can be found in the report located at: http://www.ojp.usdoj.gov/bjs/abstract/cj99.htm.

Census of Jails
(continued)

Strengths: Data are collected on a key policy issue, involvement with the justice system. There are multiple years of data available.

Limitations: This data set contains facility-level rather than individual-level data. Researchers will have counts by facility of those being confined, staff, or corrections officers who have been identified as AI/AN or NH/PI. Additional analysis related to the characteristics or experiences of these individuals is not possible using these data.

Access Requirements and Use Restrictions: Data sets are available to the public at no cost.

Contact Information: 1999 Census of Jails data can be downloaded at:
http://webapp.icpsr.umich.edu/cocoon/NACJD-STUDY/03318.xml.

Data archive information:
National Archive of Criminal Justice Data
ICPSR
University of Michigan
Institute for Social Research
P.O. Box 1248
Ann Arbor, MI 48106-1248
(800) 999-0960
(313) 763-5011
nacjd@icpsr.umich.edu
http://webapp.icpsr.umich.edu/cocoon/NACJD-SERIES/00068.xml

Questions for the Bureau of Justice Statistics should be addressed to:
James Stephan
Statistician
Bureau of Justice Statistics
810 Seventh Street, NW
Washington, DC 20531
USA
(202) 616-3289
James.Stephan@usdoj.gov
askbjs@usdoj.gov

Census of Tribal Justice Agencies in Indian Country (2002)

Sponsor:	U.S. Department of Justice/Bureau of Justice Statistics
Description:	The Census of Tribal Justice Agencies is the first comprehensive effort to identify which justice agencies operate in tribal jurisdictions, what services those agencies provide, and what information they collect and keep. The data describe the characteristics of tribal law enforcement, courts and administration, corrections and intermediate sanctions, criminal history records, and justice statistics. The data also describe the criminal justice system in Indian Country including which tribes have sworn law enforcement personnel and the source of their authority, the number and types of tribal court systems, who performs the tribal detention function and what types of sanctions are imposed, and tribal access to state and national criminal record systems.
Relevant Policy Issues:	Differences in Resolution of Arrest by Type of Court System, and Factors Contributing to Disparities in Involvement with Justice System and Outcomes.
Data Type(s):	Census survey
Unit of Analysis:	The tribal justice agency is the unit of analysis.
Identification of AI/AN/NA:	AI/AN/NA individuals are not identified in the data set.
AI/AN/NA Population in Data Set:	Ninety-two percent (314 of 341) of tribal justice agencies responded to the survey. Participation by Alaska Native communities was not sufficient to allow them to be included in the final data.
Geographic Scope:	The geographic area covered by the study is national. The data set includes 314 tribal justice agencies out of 341. The state and name of the tribe is identified for each justice agency. While the state of the agency's location is available, national analysis is recommended as there may be some states with very few agencies.
Date or Frequency:	Data were collected once.
Data Collection Methodology:	The questionnaires were distributed by mail and participants could respond by mail, fax, telephone, or online.
Participation:	Optional, without incentives.
Response Rate:	For tribal agencies, it was 92 percent. Responses were very poor for Alaska Native communities, but the rate is not reported.

Census of Tribal Justice Agencies in Indian Country (2002)
(continued)

Strengths:	The key strength of this data collection is its uniqueness. This is the only comprehensive description of the justice system in Indian Country available. The response rate for AI tribes was very high (92 percent). The data are available online in spreadsheet format for additional analysis.
Limitations:	The key weakness was the very poor response rate by AN communities. Response was so low that they could not be included in the final reported data.
Access Requirements and Use Restrictions:	The data set is available to the public at no cost.
Contact Information:	The data and reports can be downloaded at: http://www.ojp.usdoj.gov/bjs/abstract/ctjaic02.htm. Questions for the Bureau of Justice Statistics should be addressed to: Steven W. Perry, Statistician Bureau of Justice Statistics 810 Seventh Street, NW Washington, DC 20531 USA (202) 307-0765 askbjs@usdoj.gov
Reports of Interest:	Location for final report: http://www.ojp.usdoj.gov/bjs/abstract/ctjaic02.htm.

Consumer Expenditure Surveys (CE) Interview and Diary Surveys

Sponsor: U.S. Department of Labor (DoL)/Bureau of Labor Statistics (BLS)

Description: The Consumer Expenditure Survey (CE) consists of two surveys—the quarterly Interview survey and the Diary survey—that provide information on the buying habits of American consumers, including data on their expenditures, income, and consumer unit (families and single consumers) characteristics. The Diary survey asks consumers to track their expenditures over a two-week period. The Interview survey gathers similar data in a series of quarterly computer-assisted interviews.

The CE is a basic source of data for revising the items and weights in the market basket of consumer purchases to be priced for the Consumer Price Index. It is also used to construct statistical measures of consumption, for analysis of expenditure patterns by individual and family characteristics, in market research studies, in economic research, and to develop consumer guidance materials.

Relevant Policy Issues: Measurement of Economic/Employment Disparities, Income Status, Unemployment Rates, and Economic Assistance Program Participation Rates.

Data Type(s): Survey

Unit of Analysis: Consumer Unit. A consumer unit consists of any of the following: (1) all members of a particular household who are related by blood, marriage, adoption, or other legal arrangements; (2) a person living alone or sharing a household with others or living as a roomer in a private home or lodging house or in permanent living quarters in a hotel or motel, but who is financially independent; or (3) two or more persons living together who use their incomes to make joint expenditure decisions. The terms consumer unit, family, and household are often used interchangeably for convenience. However, the proper technical term for purposes of the Consumer Expenditure Survey is consumer unit.

Identification of AI/AN/NA: Race is self-reported. There are no specific instructions given for self-identification. Respondents are permitted to check all that apply. The categories available:

- White
- Black or African American
- American Indian or Alaska Native (AI/AN)
- Asian
- Hawaiian or Other Pacific Islander (NH/PI)

AI/AN/NA Population in Data Set: The (unweighted) numbers of AI/AN/NA represented in the CE Diary 2004 are:

TOTAL: 14,917
AI/AN: 80
NH/PI: 55

The (unweighted) numbers of AI/AN/NA represented in the CE Interview 2004 are:

TOTAL: 38,844
AI/AN: 204
NH/PI: 118

Geographic Scope: The geographic scope of the CE is national. The microdata includes the following geographic identifiers that would support analyses: some states, city population size, and rural vs. urban areas.

The Bureau of Labor Statistics makes state identifiers available for use with the public-use CD-ROMs, although some states are not identified because the sample size for each state is very small. Further information about geographic identifiers can be obtained from the Bureau's Division of Consumer Expenditure Surveys, by e-mail at cexinfo@bls.gov or by telephone at (202) 691-6900.

Consumer expenditure data for selected Metropolitan Statistical Areas (MSAs) also are published in the biennial reports. For confidentiality reasons, MSA identifiers are not included on the public-use microdata on CD-ROMs.

Date or Frequency: The CE has been conducted annually since 1984 and will continue to be conducted annually. Data are available for all years the CE has been conducted.

Aggregation: BLS sells microdata for the 1984-2004 CE on CD-ROM. These data can be integrated if careful attention is paid to differences across the forms from collection to collection as well as differences in sampling design from collection to collection. BLS recommends that to represent the covered population in aggregating the data, the quarterly weights should be used for each consumer unit. See the following link for what is available and price lists: http://www.bls.gov/cex/csxmicro.htm

If the 2003 and 2004 data were aggregated, the resulting (unweighted) Ns would be:

<u>Diary data</u>
TOTAL: 30,744
AI/AN: 176
NH/PI: 109

<u>Interview data</u>
TOTAL: 79,218
AI/AN: 417
NH/PI: 196

Consumer Expenditure Surveys (CE) Interview and Diary Surveys (continued)

Data Collection Methodology: The diary component uses a self-administered paper-and-pencil diary that is returned to BLS. The interview component is conducted in person using a computer-assisted personal interview protocol or a laptop computer.

Participation: Optional, without incentives

Response Rate: The following table represents response rate information for the 2001, 2002 and 2003 CE:

	Interview	Diary
2001- Percent of eligible units interviewed	78.0	74.9
2002- Percent of eligible units interviewed	78.0	74.1
2003- Percent of eligible units interviewed	78.6	73.4

Sampling Methodology: The CE is a national probability sample of households designed to represent the total U.S. civilian noninstitutional population. The selection of households begins with the definition and selection of primary sampling units (PSUs), which consist of counties (or parts thereof), groups of counties, or independent cities. Within these PSUs, the sampling frame (that is, the list from which housing units are chosen) for the Consumer Expenditure Survey is generated from the 2000 Census 100-percent detail file. The frame is augmented by a sample drawn from new construction permits and extra housing units identified through improvements in coverage techniques.

The Census Bureau selects a sample of approximately 12,500 addresses for participation in the Diary survey each year. The Interview survey is a rotating panel survey in which approximately 15,000 addresses are contacted in each calendar quarter of the year. One-fifth of the addresses contacted each quarter are new to the survey and provide the bounding interviews that afford baseline data, but are not used to compute the survey's published expenditure estimates. After a housing unit has been in the sample for five consecutive quarters, it is dropped from the panel and a new housing unit is selected to replace it.

Analysis: Beginning with year 2000 data, the Consumer Expenditure Survey program made available standard error tables using integrated data from both surveys. These standard error tables correspond to the program's standard tables, except for the classification by region, population size of area of residence, and selected age.

Selected standard error tables are available at http://www.bls.gov/cex/csxstnderror.htm.

Detailed information on how to use the standard error tables is provided at http://www.bls.gov/cex/anthology/csxanth5.pdf.

Authorization: The Bureau of Labor Statistics conducts the Consumer Expenditure Diary Survey under the authority of Title 29 of the U.S. Code. Congress authorizes the financial support for the survey through Public Laws 94-439 and 95-205.

Consumer Expenditure Surveys (CE) Interview and Diary Surveys (continued)

Strengths: The CE is a very well-documented series of studies; much of this documentation is available on-line. Multiple years of data are available.

Limitations: There are a very small number of AI/AN/NA respondents in the data from the CE Diary 2004 and only a moderate number in the CE Interview 2004. For maximum utility, CE data should be aggregated across multiple years to increase the numbers of AI/AN/NA respondents represented in the sample. Aggregation of the data, however, may require sophisticated programming and statistical skills.

Access Requirements and Use Restrictions: Microdata are available to the public. Cost per annual CD is $145.

Contact Information:

Bureau of Labor Statistics
Consumer Expenditure Surveys -- Branch of Information and Analysis
Postal Square Building, Room 3985
2 Massachusetts Avenue, N.E.
Washington, DC 20212-0001

Telephone- (202) 691-6900
FAX- (202) 691-7006
E Mail- CEXINFO@bls.gov

For information on how to obtain the actual data, please contact BLS officials using the contact information given above. There is also an online table generator available at the CE homepage at: http://www.bls.gov/cex/.

Current Population Survey (CPS)

Sponsor: U.S. Department of Labor (DoL)/Bureau of Labor Statistics and U.S. Department of Commerce/Bureau of the Census

Description: The Current Population Survey (CPS) is the primary source of information on the labor force characteristics of the U.S. population. The sample is selected to represent the civilian noninstitutional population. Respondents are interviewed to obtain information about the employment status of each member of the household 15 years of age and older. Data collected include employment; unemployment; earnings; hours of work; a variety of demographic characteristics including age, sex, race, marital status, and educational attainment; occupation, industry, and class of worker. Supplemental questions are often asked on a variety of topics including school enrollment, income, previous work experience, health, employee benefits, and work schedules.

Relevant Policy Issues: Income Status, Unemployment Rates, Economic Opportunity, and Demographic and Economic Indicators.

Data Type(s): Survey

Unit of Analysis: Individual

Identification of AI/AN/NA: Participants were asked to respond to the question on race by indicating one or more of six race categories. The six race categories are:

- White
- Black or African American
- American Indian/Alaska Native (AI/AN)
- Asian
- Native Hawaiian/Other Pacific Islander (NH/PI)
- Some Other Race (this category is not read/displayed to the respondent)

Responses to the race item are recoded into multiple race categories for analytic purposes. Those categories that include AI/AN or NH/PI are listed below:

- AI/AN Only
- NH/PI Only
- White/AI/AN
- White/NH/PI
- Black/AI/AN
- Black/NH/PI
- AI/AN/Asian
- Asian/NH/PI
- White/Black/AI/AN
- White/AI/AN/Asian
- White/Asian/NH/PI
- White/Black/AI/AN/Asian

AI/AN/NA Population in Data Set:

Responses to the race item were recoded into the multiple race categories. The following categories reflect the unweighted counts for AI/AN/NA respondents in the April, March, and February 2006 CPS:

February 2006 (N = 136,294)
AI/AN Only: 1,510
NH/PI Only: 462
White/AI/AN: 1,357
White/NH/PI: 149
Black/AI/AN: 143
Black/NH/PI: 22
AI/AN/Asian: 4
Asian/NH/PI: 124
White/Black/AI/AN: 102
White/AI/AN/Asian: 17
White/Asian/NH/PI: 156
White/Black/AI/AN/Asian: 5

March 2006 (N = 135,028)
AI/AN Only: 1,447
NH/PI Only: 472
White/AI/AN: 1,345
White/NH/PI: 155
Black/AI/AN: 146
Black/NH/PI: 23
AI/AN/Asian: 6
Asian/NH/PI: 128
White/Black/AI/AN: 95
White/AI/AN/Asian: 19
White/Asian/NH/PI: 171
White/Black/AI/AN/Asian: 3

April 2006 (N = 136,405)
AI/AN Only: 1,486
NH/PI Only: 486
White/AI/AN: 1,351
White/NH/PI: 165
Black/AI/AN: 150
Black/NH/PI: 20
AI/AN/Asian: 6
Asian/NH/PI: 110
White/Black/AI/AN: 91
White/AI/AN/Asian: 21
White/Asian/NH/PI: 188
White/Black/AI/AN/Asian: 4

Geographic Scope:

The geographic scope of the study is national.

There are several geographic variables on the data sets that could be useful for analysis. They include:

- Region (Northeast, Midwest, South, and West);
- State Census Code (state names);
- Combined Statistical Area Federal Information Processing Standards Code (FIPS) (e.g., Appleton-Oshkosh-Neenah,WI; Chicago-Naperville-Michigan City, IL- IN-WI; Cincinnati-Middletown-Wilmington, OH-KY-IN; Cleveland-Akron-Elyria, OH);
- Metropolitan Statistical Area FIPS Code (e.g., Appleton-Oshkosh-Neenah, WI, MSA; Grand Rapids-Muskegon-Holland, MI MSA; Greenville-Spartanburg-Anderson, SC MSA);
- Principal City/Balance Status (Principal City, Balance Metropolitan, Nonmetropolitan, Not Identified);
- Metropolitan Status (Metropolitan, Nonmetropolitan, Not Identified);
- Individual Central City Code (specific city code); and
- Metropolitan Statistical Area Size (100,000 - 249,999, 250,000 - 499,999, 500,000 - 999,999, 1,000,000 - 2,499,999, 2,500,000 - 4,999,999, 5,000,000+).

Date or Frequency:

The CPS is a rotating panel survey that has been conducted monthly for over 50 years. A panel survey is a survey in which similar measurements are made on the same sample at different points in time, and in a rotating panel survey, part of the sample is changed each month. For the CPS, each monthly sample is divided into eight representative subsamples or rotation groups. A given rotation group is interviewed for a total of 8 months, divided into two equal periods. It is in the sample for 4 consecutive months, leaves the sample during the following 8 months, and then returns for another 4 consecutive months. In each monthly sample, one of the eight rotation groups is in the first month of enumeration, another rotation group is in the second month, and so on. Under this system, 75 percent of the sample is common from month to month and 50 percent is common from year to year for the same month. This procedure provides a substantial amount of month-to-month and year-to-year overlap in the sample, thus providing better estimates of change and reducing discontinuities in the data series without burdening any specific group of households with an unduly long period of inquiry.

Aggregation:

Public release cross-sectional data are available for each month of data collection and smaller scope cross-wave tables of data are also publicly available. It is possible to combine the cross-sectional data sets to obtain cross-wave data sets that would contain information not available in the cross-wave data sets that are currently provided. The overall count of members of the AI/AN/NA population in an aggregated data set, however, would not increase dramatically as the CPS is a panel survey with about 75 percent overlap between samples from month to month.

Current Population Survey (CPS)
(continued)

Data Collection Methodology:	The mode of data collection for the CPS is both telephone and in-person interviewing.
Participation:	Optional, without incentives
Response Rate:	Nonresponse rates are less than 9 percent for the monthly CPS for September 2003 through September 2004.
Sampling Methodology:	The CPS sample is a multistage stratified sample of approximately 72,000 households. Of these households, approximately 56,000 housing units from 792 sample areas were interviewed. The CPS samples housing units from lists of addresses obtained from the 1990 Decennial Census of Population and Housing. These lists are updated continuously for new housing built after the 1990 census. The first stage of sampling involves dividing the United States into primary sampling units (PSUs)—most of which comprise a metropolitan area, a large county, or a group of smaller counties. Every PSU falls within the boundary of a state. The PSUs are then grouped into strata.
Analysis:	Effective sample size, design effects, and standard errors for estimates are discussed in detail in the following publication: *Technical Paper 63RV: Current Population Survey - Design and Methodology* (http://www.census.gov/prod/2002pubs/tp63rv.pdf).
Authorization:	The information collected in the CPS is authorized by the following: Title 13, U.S. Code, Section 182 (Authorizes the Census Bureau to collect statistical information); Title 29, U.S. Code, Sections 1-9 (Authorizes the Bureau of Labor Statistics to collect labor force statistics); Title 38, U.S. Code, Section 219 (Authorizes the Census Bureau to collect information for the Department of Veterans' Affairs); and Public Laws 89-10, 92-318, 93-380 (Authorizes the Census Bureau to collect information on education).
Strengths:	The CPS data sets contain a large number of AI/AN/NA respondents. They include information on key policy issues. There are multiple years of data available.
Limitations:	Aggregation of the monthly data to obtain a longitudinal data set would require the expertise of a skilled statistician.
Access Requirements and Use Restrictions:	CPS data are available to the public at no cost.

Contact Information:

Information on the CPS is available at http://www.bls.census.gov/cps/.

Email inquiries can be submitted via the "Ask a Question" tab on the ask.census.gov webpage. (https://ask.census.gov/cgi-bin/askcensus.cfg/php/enduser/std_alp.php).

Telephone inquiries can be made at (301) 763-3806.

Basic monthly data for May 2004 through April 2006 are currently available from DataFerrett. DataFerrett is a data mining tool that accesses data stored in TheDataWeb (a network of online data libraries) through the Internet. DataFerrett must be installed as an application on a personal computer or used as a java applet with an Internet browser (http://dataferrett.census.gov/). Older data are available via file transfer protocol (ftp) from the CPS website: http://www.bls.census.gov/ferretftp.htm.

Early Childhood Longitudinal Study, Birth Cohort (ECLS-B)

Sponsor: U.S. Department of Education (DoE)/National Center for Educational Statistics (NCES)

Description: The Early Childhood Longitudinal Study is designed to provide decision-makers, researchers, child care providers, teachers, and parents with detailed information about children's early life experiences. The Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) looks at children's health, development, care, and education during the formative years from birth through kindergarten entry. The ECLS-B selected a nationally representative sample of children born in the year 2001 to follow from birth through kindergarten.

Relevant Policy Issues: Educational Attainment, Measures of Well-being for Families/Households, and Measures of Well-being for Children.

Data Type(s): Survey

Unit of Analysis: Individual

Identification of AI/AN/NA: The ECLS-B collects information on race and ethnicity in two places: the parent interview and the birth certificate. Race/ethnicity information from the birth certificate was used for sampling purposes only. For analytic purposes, ECLS-B recommends using the information provided in the parent interview.

In the parent interview, children’s race/ethnicity is defined by a series of variables. Parents were asked whether their child was of Spanish, Hispanic, or Latino origin. The parents were then shown a card with race response options and asked to choose from a number of options.

The restricted-use ECLS-B data files identify the following race categories:

- White
- Black
- American Indian
- Asian Indian
- Chinese
- Filipino
- Japanese
- Korean
- Vietnamese
- Other Asian
- Native Hawaiian
- Guamanian
- Chamorro
- Samoan
- Other Pacific Islander

The data set allows for children to be identified as more than one race.

In the second data collection wave, parents of children who were identified as AI/AN in the first data collection wave were asked to confirm that they/their children were AI/AN. If confirmed, interviewers asked "[Are you/Is [the child]] formally enrolled in that (tribe/Alaska Regional Corporation)?" and "[Do you/Does [the child]] currently live on tribal lands or a reservation?"

AI/AN/NA Population in Data Set:

During the first wave of the study, parents of approximately 10,700 children completed interviews, and approximately 10,200 children were directly assessed. The count of AI/AN individuals in the 2001-2002 base year of the ECLS-B study (data were first collected when the children were approximately 9 months old) are provided below:

Total AI/AN population = 750*
AI/AN and Hispanic = 150*
AI/AN, non-Hispanic = 300*
AI/AN, non-Hispanic, more than one race = 300*

When appropriately weighted to be nationally representative, this sample of AI/AN children represents approximately 2 percent of all children born in the United States in 2001.

For the second year (children approximately 2 years old) collected in 2003:
Total AI/AN population = 700*

*Please note: These counts have been rounded according to NCES rounding rules, as the ECLS-B data are currently only available in a restricted format.

Although "Native Hawaiian" and "Other Pacific Islander" are provided as separate racial categories in the data files, reports present data for these groups rolled up into a single category: "Other Asian/Pacific Islander."

AI/AN/NA Subpopulations:

The restricted-use data files contain all the detailed race/ethnicity information colleted in the parent interview. Relevant subpopulations in the ECLS-B data include Native Hawaiians, Guamanians, Chamorros, Samoans, and Other Pacific Islanders. The restricted-use data file also includes information about whether the child is formally affiliated with a tribe and lives on a reservation.

Geographic Scope:

The geographical scope of the ECLS-B is national. The sample is designed also to support regional estimates. It is not designed to estimate characteristics at the state level.

Date or Frequency:

Wave 1: Data collection in the first wave took place between fall 2001 and fall 2002, at which time most of the sampled children were about 9 months of age (65 percent of AIAN children were 8 to 10 months of age).

Wave 2: Children were about 2-year-olds (collected in 2003)

Wave 3: Children were preschool-aged (e.g., age 4) (collected in 2005)

Wave 4: Children will be in kindergarten (to be collected 2006-2007)

Wave 5: Includes children who were not yet enrolled in kindergarten during the Wave 4 field period (to be collected in fall 2007)

Data Collection Methodology: ECLS-B data collection in the first wave (when the children were about 9 months old) involved three parts:

Child assessment: Children participated in a variety of activities to assess their early mental, physical, and socioemotional development. A trained staff member measured the child's mental and physical skills through an untimed one-on-one assessment in the child's home. Assessment tools included the Bayley Short Form-Research Edition (BSF-R) (a variation of the Bayley Scales of Infant Development-Second Edition that was developed specifically for use in the ECLS-B), the Nursing Child Assessment Teaching Scale (NCATS), and physical measurements.

Parent interview: Parents/guardians were asked to provide key information about their children and themselves. The parent interview included two instruments: the parent interview instrument and the parent self-administered questionnaire (PSAQ). The first was conducted in person by trained field interviewers using computer-assisted personal interviewing (CAPI) as part of the home visit. The PSAQ was a paper and pencil instrument, presented during the parent CAPI instrument for the respondent to complete and return in a provided envelope, and contained 23 questions on topics some people might prefer to answer privately.

Father questionnaires: The ECLS–B also collected data from fathers through two separate father questionnaires: the resident father questionnaire and the nonresident father questionnaire. The nonresident father questionnaire was only administered in cases where the child did not live in the same household as his or her biological father and a minimum contact frequency was met. Both father questionnaires were self-administered with telephone follow-up.

Participation: Optional, with incentives. For the first round, parent participants received $50 and a book for their child. For the second round, parents received $30 and a children's book.

Response Rate: NCES reported the response rate for Wave 1 as 74.1 percent overall while the AI/AN response rate was reported as 79.3 percent (based on weighted data). The response rate is calculated as the weighted number of completed parent interviews divided by the total eligible sample. To be considered complete, the first three sections of the parent interview needed to be completed.

Early Childhood Longitudinal Study, Birth Cohort (ECLS-B)
(continued)

Sampling Methodology:

The sample for ECLS-B was selected using a clustered, list frame sampling design. The list frame was registered births in the National Center for Health Statistics' (NCHS) vital statistics system (from lists provided by state registrars). Births were sampled from 96 core primary sampling units (PSUs) representing all infants born in the United States in the year 2001. The PSUs were counties and county groups.

Sampling was based on occurrence of the birth as listed on the birth certificate. Sampled children subsequently identified by the state registrars as having died or who had been adopted after the issuance of the birth certificate were excluded from the sample. Also, infants whose birth mothers were younger than 15 years at the time of the child's birth were excluded.

Oversample of AI/AN/NA Population:

Eighteen additional PSUs were selected from a supplemental frame consisting of areas where the population has a higher proportion of AI/AN births. The PSUs in the AI/AN PSU sampling frame were counties or groups of counties that had at least an expected 50 AI/AN sample births based on 1994-1996 National Center for Health Statistics natality detail files and that had relatively large proportions of AI/AN births.

Analysis:

The effective sample size based on the number of complete cases in wave 1 for the AI/AN population is 1,190.
Design effects (weighting effect) = 1.0500

Strengths:

Documentation is extremely detailed and very clear. The study includes some oversamples of American Indian/Native Americans. An extensive nonresponse bias analysis was conducted, and findings from these analyses suggest that there is not a bias due to nonresponse. Details on the nonresponse bias analysis are available in the study's documentation.

Limitations:

The Institutional Review Board of the Navajo Nation reservations did not approve participation in the study. Where cases were drawn from persons residing on a Navajo Nation reservation, those cases were treated as nonresponse. Navajos not living on reservations were included in the sample.

Access Requirements and Use Restrictions:

The data are available to researchers with an NCES restricted-use license. The steps for obtaining a license are detailed here:
http://nces.ed.gov/pubsearch/licenses.asp.

Contact Information:

The ECLS-B staff can be contacted by sending an email to: ecls@ed.gov.

Questions about NCES restricted-use licenses can be addressed to:
Cynthia L. Barton
Data Security Assistant
Phone: (202) 502-7307
E-mail: Cynthia.Barton@ed.gov

Early Childhood Longitudinal Study, Birth Cohort (ECLS-B)
(continued)

Reports of Interest:	Flanagan, K., and Park, J. (2005). *American Indian and Alaska Native Children: Findings From the Base Year of the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B)* (NCES 2005–116). U.S. Department of Education. Washington, DC: National Center for Education Statistics

Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K)

Sponsor: U.S. Department of Education (DoE)/National Center for Educational Statistics (NCES)

Description: The Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K) is an ongoing study that focuses on children's early school experiences beginning with kindergarten and following children through 12th grade. In the fall of 1998, ECLS-K began following a nationally representative sample of kindergarteners. The ECLS-K provides descriptive information on children's status at entry to school, their transition into school, and their progression through 8th grade. (Initially, the ECLS-K was designed to follow children through their fifth grade year in school – however plans have been made to extend the study to follow the ECLS-K children through their eighth grade of school. The study will end with the data collection scheduled for school year 2006-2007.)

The longitudinal nature of the ECLS-K data enables researchers to study how a wide range of family, school, community, and individual factors are associated with school performance. Researchers can request the child-level files for each year of data collection, as well as the longitudinal kindergarten to fifth grade data file. Data are collected from a direct child assessment, from parent interviews, from school administrators and teachers, and from student records and a school facilities checklist.

Relevant Policy Issues: Educational Attainment, Educational Opportunities, Factors Contributing to Educational Disparities, and Identification of Evidence-based Practices and Programs that Produce Positive Educational Outcomes and are Generalizable/replicable.

Data Type(s): Survey

Unit of Analysis: Individual child

Identification of AI/AN/NA: Race information is obtained during the parent interviews using the following question: What is your race? (The same question is asked concerning the child: "What is [NAME OF CHILD'S] race?")

Categories include:

- American Indian or Alaska Native (AI/AN)
- Asian
- Black or African American
- Native Hawaiian or other Pacific Islander (NH/PI)
- White
- Another race (specify)

When the parent interview is not completed, race information is obtained from school records. The parent interview is considered the best source of information regarding race/ethnicity.

Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K) (continued)

AI/AN/NA Population in Data Set:

Completed interviews with AI/AN/NA children, parents, and school informants by type of data collection for the school year 2003-2004 (fifth-grade):

Child assessment
Total completed interviews: 11,260
AI/AN: 210
NH/PI: 144

Parent interview
Total completed interviews: 10,913
AI/AN: 222
NH/PI: 136

School administrator questionnaires
Total completed interviews: 10,937
AI/AN: 191
NH/PI: 145

School facilities checklist
Total completed interviews: 11,154
AI/AN: 208
NH/PI: 146

Student records abstract
Total completed interviews: 10,015
AI/AN: 197
NH/PI: 125

Teacher-level questionnaires
Total completed interviews: 10,872
AI/AN: 206
NH/PI: 138

Geographic Scope:

The geographic scope of the study is national. The ECLS-K public-use data also contain information on the regional location of the child's school (i.e., Northeast, Midwest, South, and West.) The ECLS-K sample was designed to support national and regional estimates. It was not designed to estimate characteristics of children, teachers, families and schools at the state level. Variables such as the child's home and school zip code are suppressed on the public use files to ensure respondent confidentiality.

Date or Frequency:

The ECLS-K is a longitudinal study. The same children are followed periodically from kindergarten through the 8th grade. Information was collected in the fall and the spring of kindergarten (1998-99), the fall and spring of first grade (1999-2000), the spring of third grade (2002) and the spring of 5th (2004). Future data collections will include 8th grade (2007).

Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K) (continued)

Data Collection Methodology: To collect information from children, a trained assessor visits the children in their schools. Children are assessed, un-timed, one-on-one within their school. The direct child assessment collects information about children's reading and mathematics skills, their general knowledge (i.e., science and social studies) in kindergarten and first grade, and their science knowledge in third and fifth grade. In addition, the assessment includes height/weight measurements, and in fall kindergarten only, children's psychomotor skills (e.g., hopping, skipping, jumping, manipulating blocks, drawing figures) are assessed. The direct child assessments are administered using computer assisted personal interviews (CAPI).

To collect information from parents, a trained interviewer phones the parent at their home and administers a 45-50 minute interview. Computer assisted telephone interviewing (CATI) methods are used to record the parent's answers. If the child's family does not have a telephone, the interview is conducted in person.

To collect information from schools, teachers and school administrators complete paper and pencil surveys and retrieve information from school records.

Participation: Optional, with incentives

Response Rate: Overall unweighted response rates for the fifth grade cohort (school year 2003-2004):
Child assessments: 93.6%
Parent interview: 90.7%
School administrator questionnaire: 89.6%
Facilities checklist: 91.4%
Student records abstract: 82.1%
Teacher level-questionnaire: 90.6%

There are also some differences in response rates by race. For example, in the spring-fifth grade data collection, the AI/ANs had the lowest child assessment weighted completion rate (78.3 percent) and the highest parent interview weighted completion rate (95.2 percent).

Sampling Methodology: The ECLS-K study uses a multistage probability sample design. Primary sampling units (PSUs) consist of counties and groups of counties. The second stage units are schools within the sampled PSUs. In the base year of the ECLS-K, a total of 1,277 schools were selected for the sample. The third and final stage is students within the selected schools. In the base year of the ECLS-K, a total of 22,666 students were selected. The children in ECLS-K attended both public and private schools, including both full-day and part-day kindergarten programs.

Oversample of AI/AN/NA Population: Asians and Pacific Islanders were oversampled for the ECLS-K study. To create this oversample, two independent sampling strata were formed within each school. One stratum consisted of Asian/Pacific Islander children, while the second stratum consisted of the remaining children.

Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K) (continued)

Analysis: The ECLS-K data are weighted to compensate for differential probabilities of selection at each sampling stage and to adjust for the effects of nonresponse. The User's Manual provides a detailed description of how weights were calculated and how they should be applied to the data set.

Each survey item in the ECLS-K has its own design effect that can be estimated from the survey data. The median design effects, as reported in the ECLS-K 5th Grade User's Manual, for the race/ethnicity variable for all six rounds of data collection are:

- White: 2.920
- Black : 2.532
- Hispanic: 2.456
- Asian: 3.106
- Native Hawaiian or other Pacific Islander: 4.186
- American Indian or Alaska Native: 7.058
- Other: 2.423

Strengths: Data are collected on a key policy issue, education. There are multiple years of data available. The study includes some oversampling of Pacific Islanders.

Limitations: There are a limited number of AI/AN/NA in these data sets.

Other: While the ECLS-K study includes children that attend Bureau of Indian Affairs (BIA) schools, the number of children attending these schools is suppressed to protect respondent confidentiality. Additionally, attendance at BIA schools is not included as an identifier in the data files.

Access Requirements and Use Restrictions: Unlike the Early Childhood Longitudinal Study Birth Cohort data, the ECLS-K data has a public use data file. There is also a restricted ECLS-K data set available to researchers with a National Center for Education Statistics (NCES) license. The restricted data set contains a few values and variables that are suppressed on the public use data set. For most research, the ECLS-K public use data set should suffice. For researchers who feel they need the restricted use data set, the steps for obtaining an NCES license are detailed here: http://nces.ed.gov/pubsearch/licenses.asp. There is no cost associated with use of the data.

Contact Information: Public use data file can be ordered on CD-ROM from www.edpubs.org.
Elvira Germino Hausken
Project Officer, ECLS-K
U.S. Department of Education
National Center for Education Statistics
(202) 502-7352
ECLS@ed.gov
Web Site: http://nces.ed.gov/ecls/

Food Stamp Program Quality Control Database (FSPQC)

Sponsor: U.S. Department of Agriculture (USDA)

Description: The Food Stamp Program Quality Control Database contains detailed demographic, economic, and Food Stamp Program (FSP) eligibility information for a nationally representative sample of approximately 50,000 participating households. The FSPQC data are generated from monthly quality control (QC) reviews of FSP cases that are conducted by state FSP agencies to assess the accuracy of eligibility determinations and benefit calculations for the state's FSP caseload. These data, which are produced annually, are suitable for tabulations of characteristics of food stamp units and for simulating the impact of various FSP policy changes on households and persons currently receiving food stamps. The FSPQC Database is an edited version of the raw datafile generated by the Food Stamp Program's Quality Control (QC) System.

Relevant Policy Issues: Economic Assistance Program Participation Rates, Measures of Well-being for Families/Households, and Factors Contributing to Well-being Disparities of Families.

Data Type(s): Program enrollment data

Unit of Analysis: Analysis is possible at the individual and household levels. The FSPQC lists characteristics for 1-16 members of a household. Individuals included in the database as household members are those eligible for participation in the FSP as well as those who would be considered part of the FSP household but are ineligible to participate because of a variety of reasons. In many of these cases, the income of an ineligible household member is factored into determining the benefit for the eligible portion of the household. Not included in this database are those individuals that are living under "the same roof" who can be considered a separate FSP household. Examples of these are the elderly or disabled (who have special household status rules) as well as unrelated housemates who purchase and prepare their meals separately.

Identification of AI/AN/NA: Race/ethnicity is self-reported on the application for Food Stamp benefits. The QC reviewer takes this information and incorporates it into the FSPQC database. The reporting categories in the current database are:

- White, not of Hispanic origin
- Black, not of Hispanic origin
- Hispanic
- Asian or Pacific Islander
- American Indian or Alaskan Native (AI/AN)

USDA recently announced that these categories are being changed. States have until April 1, 2007 to implement this change. The new categories will capture the following:

Food Stamp Program Quality Control Database (FSPQC)
(continued)

- American Indian or Alaskan Native
- Asian
- Black or African-American
- Native Hawaiian or other Pacific Islander
- White
- American Indian or Alaska Native and White
- Asian and White
- Black or African American and White
- American Indian or Alaska Native and Black or African-American

Additionally, separate categories will be included for each of the race categories above to designate Hispanic origin (e.g., White/Hispanic, White/Non-Hispanic.)

The FSPQC database will be revised to reflect these new reporting requirements for FY 2007.

AI/AN/NA Population in Data Set: Out of the total 48,806 household-level records in the FSPQC, there are 117,456 individuals identified across 16 person-level variables. Of these, 4,050 individuals are coded as AI/AN in 1,371 households. Of the 4,050 AI/AN individuals, 4,013 of them participate in the Food Stamp Program.

Geographic Scope: The geographic scope of the FSPQC is national. A county Federal Information Processing Standards (FIPS) code is assigned to each unit on the FSPQC file. However, the sample size does not allow analyses at the county level. The sample size is sufficient to allow analyses at the state level.

Date or Frequency: The FSPQC is compiled on a yearly basis. The most recent version of the FSPQC available to researchers is from FY 2004. The FSPQC data are typically released in late summer or early fall.

Data Collection Methodology: State FSP agencies conduct monthly case reviews to assess the accuracy of eligibility determinations and benefit calculations for the state's FSP caseload. The public use FSPQC database contains all case reviews except those removed from the file because there is too little information. These include those coded as not subject to review, those whose review was incomplete, those who are ineligible, and a few households dropped due to inconsistencies in the data.

Participation: Mandatory. States must report data to Food and Nutrition Service (FNS).

Sampling Methodology: All state agencies (including the District of Columbia, Guam and the Virgin Islands) are required to select monthly a statistically random sample from a universe of all households receiving Food Stamp benefits for that given month. Most state agencies draw the samples systematically (i.e. using a constant sampling interval), though there are some state agencies which employ simple random and/or stratified sampling techniques. All sampling plans must be approved by FNS. Required annual sample sizes range from 300 for state agencies with small Food Stamp caseloads (e.g. Wyoming and Guam), to over

1000 for larger states, with the average being around 950 per state. State agencies are required to complete reviews for at least 98 percent of those selected cases which are deemed to be part of the desired Food Stamp universe. The review findings and data for each state are reported to FNS when the review is completed. These data form the basis for the public FSPQC database.

Strengths: The FSPQC database contains a large number of AI/AN/NA respondents. The data are collected on key policy issues, including family well-being. There are multiple years of data available with little missing data (less than 1 percent missing data for the race variable in 48 states). The documentation is very detailed.

Limitations: The Food Distribution Program on Indian Reservations (FDPIR) is administered at the federal level by the FNS in cooperation with 98 tribal organizations and 6 state agencies. Many Native Americans actually participate in the FDPIR rather than the Food Stamp Program because of rural isolation and the lack of easy access to food stores. Therefore, the FSPQC underrepresents American Indians who live on reservations and receive nutrition assistance.

Additionally, the FSPQC data are limited regarding the asset and vehicle holdings of FSP participants, and there are no data available for eligible non-participants. There are also no data available for those receiving disaster benefits, as these cases are not subject to review.

Access Requirements and Use Restrictions: Data are available to the public at no cost.

Contact Information: The contact for obtaining FSPQC data is as follows:

Office of Analysis, Nutrition, and Evaluation
USDA Food and Nutrition Service
3101 Park Center Drive, Room 1014
Alexandria, Virginia 22302
703-305-2017

Additionally, a restricted version of the FSPQC data can be downloaded from the following website:
http://host4.mathematica-mpr.com/fns/fnsqcdata/download.htm.

Reports of Interest: *Background Report on the Use and Impact of Food Assistance Programs on Indian Reservations*. January 2005. Finegold, K., Pindus, N., Wherry, L., Nelson, S., Triplett, T., Capps, R. The Urban Institute.

Characteristics of Food Stamp Households: Fiscal Year 2004. September 2005. Anni Poikolainen. Mathematica Policy Research, Inc.

Hawaii Health Survey (HHS)

Sponsor: State of Hawaii Department of Health (DOH)/Office of Health Status Monitoring (OHSM)

Description: The Hawaii Health Survey (HHS) is a continuous statewide household survey of health and socio-demographic conditions. The HHS was modeled after the National Health Information Survey (NHIS). The HHS was initiated in 1968, and in 1996 it became a telephone survey. The survey is conducted for the purpose of providing Hawaii Department of Health (DOH) programs, other agencies, and the public with statistics for planning and evaluation of health services, programs, and problems. Sample data are adjusted and weighted to generate estimates of the population in Hawaii. The survey provides demographic information for observing population changes during the intercensal decade. The survey provides information on health and demographic characteristics of the people of Hawaii (e.g., income, race, education, marital status, employment, household size, insurance status, health status, morbidity, food security, and physical and mental health). However, many of the items unrelated to health are added by private agencies. For this reason, the Hawaii DOH can only run customized data analysis on the core survey items related to health issues.

Relevant Policy Issues: Measurement of Health Status.

Data Type(s): Survey

Unit of Analysis: The unit of analysis is the household or family unit, with each household's individuals identified as separate data elements. A household is defined as all persons who occupy a housing unit (e.g., house, apartment), whether or not they are related to each other.

Identification of AI/AN/NA: Race is self-reported by the respondent. Race and ethnicity are gathered according to two independent procedures in the HHS file. First, a set of variables is used to collect respondent-reported ethnicity of the parents of each household member. The respondent can list up to four ethnic categories for each member's mother and father, resulting in up to eight indicators of ethnicity for household members. Response categories include:

- White/Caucasian
- Hawaiian
- Chinese
- Filipino
- Japanese
- Korean
- Vietnamese
- Asian Indian
- Other Asian
- Samoan/Tongan
- Black/African American

- Native American/Aleut/Eskimo/Inuit
- Puerto Rican
- Mexican
- Portuguese
- Guamanian/Chamorro
- Other Pacific Islander

Additionally, the interviewer can mark an "other" category and record an ethnic background not included in the above list. Interviewers can also record responses of "don't know" (respondent may not know all ethnic background classifications of all household members) or "refused." Data are stored in the file as eight or more (8 responses and also other) ethnic indicators for every household member, and can support classification of ethnicity or race according to any classification scheme including single and multiple class schemes.

OHSM then codes the eight possible choices for each individual to one ethnic indicator based on parents' race/ethnicity according to a system consistent with U.S. Census rules for coding race/ethnicity. Specifically, if Hawaiian is listed for either the mother or father, ethnicity for that person is coded as Hawaiian.

The Hawaii Department of Health publishes figures for the following race groups: Caucasian, Hawaiian, Chinese, Filipino, Japanese, and Other.

AI/AN/NA Population in Data Set: In 2004, 6,769 households (6,769 respondents who were aged 18 years of age or older answered questions about their household and its members) were surveyed. The total number of household members described was 19,699. A breakdown of unweighted respondent groups by race is not available.

AI/AN/NA Subpopulations: Native Hawaiian (NH) alone is identified.

Geographic Scope: The geographic scope of the study is the state of Hawaii. Geographic indicators include county, island, zip code as reported by the respondent, and telephone prefix. Geographic analysis by any of these variables would not be appropriate given the small sample size.

Date or Frequency: The HHS began in 1968 and is conducted on an annual basis.

Data Collection Methodology: The HHS is administered by telephone using computer assisted telephone interviewing (CATI). Data are collected on all members of sampled households.

Participation: Optional, without incentives

Response Rate: The response rate for this study is not available.

Hawaii Health Survey (HHS)
(continued)

Sampling Methodology: The target population for the HHS is all Hawaiian residents in the state of Hawaii in non-institutionalized housing units with working telephone service at the time of the survey. Results exclude residents of the island of Niihau, households without phones, the homeless, and persons in group quarters. The HHS uses a disproportionate stratified random-digit dialing (RDD) sample that randomizes selection within strata. The sample is disproportionately selected by island (slightly larger proportions of interviews are conducted on islands with smaller populations). The sample population is statistically adjusted to represent the population of Hawaii. The respondent is an adult 18 years of age or older who is knowledgeable about their household, rather than a randomized adult.

Strengths: Data sets contain information on individuals identified or coded as Native Hawaiian.

Limitations: Technical documentation is very limited. Researchers cannot obtain the data file, and can only request analysis on the core survey items related to health policy issues.

Access Requirements and Use Restrictions: OHSM will run customized data analysis for researchers upon request (there may be a charge depending on the nature of the request). Additionally, some data items are added on to the HHS by different agencies. These data are not available through OHSM. Researchers would need to contact the different agencies regarding the availability of these additional items.

Contact Information: Office of Health Status Monitoring
Hawaii Department of Health
1250 Punchbowl Street, Room 104
Honolulu, Hawaii 96813
Phone (808) 586-4600
Fax (808) 586-4606

Head Start Program Information Report

Sponsor: U.S. Department of Health and Human Services (DHHS)/Administration for Children and Families (ACF)/Office of Head Start

Description: The Head Start Program Information Report (PIR) collects comprehensive data on the services, staff, children and families served by more than 2,700 Head Start and Early Head Start (EHS) programs nationwide (including American Indian and Alaska Native Head Start Programs.) All programs (grantees and delegates) are required to submit a PIR for each year in which they provide services to children and families. The PIR is the primary source of programmatic data for the Head Start community, their partners, Congress, and the general public.

Staffing, enrollment, and service trend information is collected through the PIR and compiled each year for use at federal, regional, and local levels. The PIR enrollment report describes the program options provided by Head Start and Early Head Start programs and provides demographic information on the children and pregnant women served. Additional information collected in the PIR enrollment report includes funded and actual enrollment, eligibility, and turnover of enrollees. The PIR family services report provides information on Head Start and Early Head Start family characteristics, including the number and types of families served, employment status, education level, and the types of services the programs provide in response to family needs.

Relevant Policy Issues: Educational Opportunities and Child/Family Well-being.

Data Type(s): Program reporting data

Unit of Analysis: The unit of analysis is the individual Head Start programs. A program identifier variable allows for separate counts to be generated for AI/AN Head Start Programs.

Identification of AI/AN/NA: Head Start programs report the number of participants per race category based on enrollment records. The instructions for reporting race are as follows:

Report the total number of children (and pregnant women in EHS programs) by race:

- American Indian or Alaska Native (AI/AN). A person having origins in any of the original peoples of North, South, or Central America, and who maintains tribal affiliation or community attachment.
- Asian. A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent.
- Black or African American. A person having origins in any of the Black racial groups of Africa.
- Native Hawaiian or other Pacific Islander (NH/PI). A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.

- White. A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.
- Biracial or Multi-Racial. A person of two or more races.
- Other. A person reporting a race other than those listed above.
- Unspecified. Race not reported and/or unknown.

AI/AN/NA Population in Data Set:
Enrollment for Program Year 2004-2005 (reported by 2,695 programs):
Total enrollment: 1,065,225
AI/AN: 55,733 (of these individuals, 21,161 are serviced by AI/AN Head Start Programs)
NH/PI: 8,448

Geographic Scope:
Geographic areas are identified by the location of the Head Start program. The PIR provides summary data at the national, regional, state, city and zip code levels.

Date or Frequency:
The Head Start PIR is compiled on an annual basis. The most recent available data are for the 2004-2005 program year.

Data Collection Methodology:
Head Start programs can submit data to the PIR through an online reporting application or using the desktop PIR reporting software.

Participation:
Mandatory

Authorization:
According to federal mandates, all Head Start and Early Head Start programs are required to complete a PIR on an annual basis. Head Start Performance Standards are under the authority for the final rule in sections 641(a) and (d), 642(b) and (d), 644(a) and (c), and 645(h)(2) of the Head Start Act, as amended (42 U.S.C. 9801 et. Seq.).

Strengths:
Data sets contain a large number of AI/AN/NA respondents. Researchers can disaggregate the data to look at children served by AI/AN Head Start programs. Data address the key policy issues of education and child and family well-being. There are multiple years of data available.

Limitations:
Data are only available at the aggregate level.

Access Requirements and Use Restrictions:
A request to use the data should be emailed to the Office of Head Start, where it will be reviewed and approved before data access is granted. There is no charge.

Contact Information:
Office of Head Start
Office of Program Management and Operations
370 L'Enfant Plaza SW
Washington, D.C. 20447
(202) 205-8396

Health and Retirement Study (HRS)

Sponsor:	U.S. Department of Health and Human Services (DHHS)/National Institutes of Health (NIH)/National Institute on Aging (NIA)
Description:	The University of Michigan Health and Retirement Study (HRS) is a national panel study of more than 22,000 Americans over the age of 50. Sponsored by the National Institute on Aging, the study is conducted every two years (1992-2006) and includes core interviews with the sampled respondents and proxy interviews when the sampled respondents have died. The study collects data on physical and mental health, insurance coverage, financial status, family support systems, labor market status, and retirement planning.
Relevant Policy Issues:	Measurement of Health Status, Disease-specific Measurements, Income Status, Educational Attainment, Measures of Well-being for Elders, and Factors Contributing to Well-being Disparities of Elders.
Data Type(s):	Survey
Unit of Analysis:	Individual
Identification of AI/AN/NA:	The questionnaire item on race is phrased as follows: "Do you consider yourself primarily White or Caucasian, Black or African American, American Indian, or Asian, or something else?" If the respondent indicated either American Indian or Alaska Native (even though the option "Alaska Native" is not stated in the question), then these two categories are collapsed into a single category. Asians and Pacific Islanders are also collapsed into a single category. Beginning in 2006, the HRS adopted the more exhaustive Census race item that separates American Indians, Alaska Natives, Native Hawaiians and Pacific Islanders.
AI/AN/NA Population in Data Set:	For the 1992 Core interview: Total: 12,521 American Indian/Alaska Native: 162 Counts for later waves of the HRS are available from codebooks that can be accessed at no cost from the HRS website. Since the HRS adds new refresher cohorts every six years (most recently in 2004), the count for AI/AN persons may increase, but could also decrease as a result of attrition or death. The overall numbers, however, will not change drastically.
Geographic Scope:	The geographic scope of this study is national. Geographic areas are also identified by Census region.
Date or Frequency:	The HRS is a longitudinal study. Baseline interviews were conducted in-home and face-to-face beginning in 1992 for the 1931-1941 birth cohort (and their

spouses, if married, regardless of age); and in 1998 for newly added 1924-1930 and 1942-1947 birth cohorts. The HRS includes follow-ups by telephone every second year, with proxy interviews after death. Data are publicly available for each wave of data collection. Data collection for the 2006 wave was underway at the time this catalog was prepared. Also, beginning in 2006, one half of the follow-ups will be conducted face-to-face to permit collection of biological samples and physical performance measures.

Aggregation: The HRS is a longitudinal panel survey (conducted every two years since 1992). Public release cross-sectional data are available for each year of data collection and smaller-scope cross-wave data sets are also publicly available. It is possible to combine the cross-sectional data sets to obtain cross-wave data sets that would contain information not available in the cross-wave data sets that are currently provided. The overall count of members of the AI/AN population, however, will not increase as the same respondents are represented in each wave of data collection, but it may decrease because of attrition or death.

Data Collection Methodology: Most of the interviews are done by telephone, although exceptions are made when respondents have health limitations that would make an hour-long session on the telephone difficult or impossible or when there is no telephone in the household. The preferred mode of data collection was face-to-face for the first wave of data collection and by telephone for subsequent waves.

Participation: Optional, with incentives

Response Rate: The overall unweighted response rate reported for the 2004 wave of data collection was 86.2 percent. Given the complexity of the HRS design, considerable detail on the calculation of response rates across waves is available from http://hrsonline.isr.umich.edu/intro/sho_uinfo.php?hfyle=sample_new_v2&xtyp=2#rates.

Sampling Methodology: The HRS sample is selected using a multi-stage area probability sample design. The sample includes four distinct selection stages. The primary stage of sampling involves probability proportionate to size (PPS) selection of U.S. Metropolitan Statistical Areas (MSAs) and non-MSA counties. This stage is followed by a second stage sampling of area segments (secondary sampling units or SSUs) within sampled primary sampling units (PSUs). The third stage of sample selection is preceded by a complete listing (enumeration) of all housing units (HUs) that are physically located within the bounds of the selected SSU. The third sampling stage is a systematic selection of housing units from the HU listings for the sample SSUs. The fourth and final stage in the multi-stage design is the selection of an age-eligible person within a sample HU.

Analysis: Methodological details regarding the survey design of the HRS are available from http://hrsonline.isr.umich.edu/docs/sho_refs.php?hfyle=design&xtyp=2.

Health and Retirement Study (HRS)
(continued)

Strengths: Data are collected on key policy issues including health and the well-being of elders. There are multiple years of data available. Comprehensive documentation is available. There is a low sample attrition rate.

Limitations: There are a small number of AI/AN respondents and NH/PI respondents cannot be separated from Asian respondents.

Access Requirements and Use Restrictions: Data are available to the public at no cost. Detailed race data are only available as restricted-use data. Researchers will need to obtain special permission to access these files, in order to protect confidentiality.

Contact Information: Data can be accessed at: http://hrsonline.isr.umich.edu.

Health and Retirement Study
Survey Research Center
Institute for Social Research
University of Michigan
426 Thompson Street
Ann Arbor, MI 48104
Phone: (734) 936-0314
Fax: (734) 647-1186
Email: hrsquest@isr.umich.edu

Health Behavior in School-aged Children (HBSC)

Sponsor: U.S. Department of Health and Human Services (DHHS)/National Institutes of Health (NIH)/National Institute of Child Health and Human Development (NICHD) and the Health Resources and Services Administration (HRSA) (The World Health Organization (WHO) collaborates to disseminate results.)

Description: The Health Behavior in School-aged Children (HBSC) survey is a cross-national survey intended to help researchers better understand the well-being, health behaviors and social context of children aged 11, 13 and 15, who are attending school. Specifically, the survey seeks to monitor health-risk behaviors and attitudes in youth throughout the adolescent school age years to provide background and identify targets for health promotion initiatives. In addition, the survey offers insight in the development of health attitudes and behaviors through early adolescence. Although cross-national comparisons can only be made with children aged 11, 13, and 15, the U.S. surveys also includes larger, nationally representative samples of ages 11 through 15 with an over-sampling of African-American and Hispanic children. The sample size in 2001-2002 is approximately 14,800.

Questions in the survey address type of drug use (e.g., tobacco, alcohol, marijuana, cocaine, inhalants, hallucinogens, and a number of other substances), ease of obtaining drugs, frequency of drug usage, and other health behaviors and personal history such as eating habits, family make-up, depression, stealing, fighting, bringing weapons to school, anger management, attention span at school, and opinions about school itself. The U.S. study also includes a survey of school administrators and health educators to provide contextual information about the school and health education programs in the school. The HBSC began in 1983 and has been conducted approximately every 4 years. The most recently available data for the United States are the 2001-2002 survey year data.

Relevant Policy Issues: Measurement of Health Status, Health Disparities, Income Status, Measurement of Economic/Employment Disparities between AI/AN/NA and General Population, Factors Contributing to Educational Disparities, Factors Contributing to Well-being Disparities of Families, and Measures of Well-being for Children.

Data Type(s): Survey

Unit of Analysis: Individual

Identification of AI/AN/NA: The survey question is phrased as follows: What is your race? (Mark one or more races to indicate what you consider yourself to be.)

- American Indian or Alaska Native (AI/AN)
- Asian
- Black or African American
- Native Hawaiian or other Pacific Islander (NH/PI)
- White

Health Behavior in School-aged Children (HBSC)
(continued)

AI/AN/NA Population in Data Set:	In 2001-2002 survey data: Total number of records: 14,818 AI/AN: 572 NH/PI: 116
Geographic Scope:	The survey is cross-national and country of residence is the only geographic area identified. The U.S. portion of the data set does not provide any finer geographic detail.
Date or Frequency:	The survey has been conducted approximately every 4 years beginning in 1983-1984.
Data Collection Methodology:	Data are collected via self-completed questionnaires administered to children aged 11, 13, and 15 attending sampled schools.
Participation:	Optional, without incentives
Response Rate:	Of the 548 schools were selected to participate for the 2001-2002 survey year, 340 (62.5 percent) responded yielding 18,593 eligible students. Of the eligible students (after eliminating 637 absent on day of survey, 600 not providing consent when required, 518 parents declining their child permission, and 1,620 students declining to participate), 15,245 (82 percent) completed questionnaires. Of this sample, 62 students who had missing data on a significant number of key items and 365 students who were outliers (+1 percent) for age in grade were dropped, leaving a sample of 14,818 for further analyses.
Sampling Methodology:	The HBSC utilizes a three-stage cluster design where school counties are the primary sampling unit (PSU) or first stage, schools are the second stage, and classrooms are the third stage.
Analysis:	A SAS macro file, made available with the downloaded data set, enables the end user to calculate appropriate standard errors that adjust for design effects.
Strengths:	The 1997-1998 survey data include a large number of AI/AN/NA respondents.
Limitations:	The 58 percent school participation rate may impact the external validity (generalizability) of study findings to the overall population of school-aged children aged 11-17 in the U.S.
Access Requirements and Use Restrictions:	The data set is available to the public at no cost.

Health Behavior in School-aged Children (HBSC)
(continued)

Contact Information:	The data for the United States and documentation can be downloaded through the University of Michigan's Inter-University Consortium for Political and Social Research: http://webapp.icpsr.umich.edu/cocoon/SAMHDA-SERIES/00195.xml.
Reports of Interest:	Summary Report Download: http://www.euro.who.int/Document/e82923.pdf. To obtain research protocol: http://www.hbsc.org/publications/research_protocols.html.

Health Information National Trends Survey (HINTS)

Sponsor:	U.S. Department of Health and Human Services (DHHS)/National Cancer Institute (NCI)/Health Communication and Informatics Research Branch (HCIRB)/Division of Cancer Control and Population Sciences (DCCPS)
Description:	The Health Information National Trends Survey (HINTS) data collection program was created to monitor changes in the rapidly evolving field of health communication. Questions on the HINTS survey include topics such as health communication with doctors, obtaining information from the media, knowledge of cancer and screening behavior, primary cancer risk behaviors, and respondent characteristics. HINTS data were collected in 2003 and 2005. Uses of the data include: (a) extending cancer communication research from the laboratory to the population, (b) monitoring the population's use of new media (e.g., and specifically the Internet), (c) documenting the public's progress in accurate knowledge related to cancer and chronic disease prevention, and (d) stimulating cross-branch cooperation.
Relevant Policy Issues:	Identification of Evidence-based Practices and Programs that Address Causes of Health Disparities, Result in Positive Health Outcomes, and are Generalizable/replicable.
Data Type(s):	Survey
Unit of Analysis:	Individual
Identification of AI/AN/NA:	Which one or more of the following would you say is your race? Are you American Indian or Alaska Native (AI/AN), Asian, Black or African American, Native Hawaiian or other Pacific Islander (NH/PI), or White? (Interviewers are instructed to code all that apply. If the respondent says "Hispanic," they are to probe for one of the listed race categories.)
AI/AN/NA Population in Data Set:	In 2003, data were collected from 6,369 respondents. Of these 6,369 respondents, 203 were AI/AN and 32 were NH/PI. In 2005, data were collected from 5,586 respondents. Of these 5,586 respondents, 141 were AI/AN and 17 were NH/PI.
Geographic Scope:	The geographic scope of the study is national. No additional geographic breakdowns of the data are available.
Date or Frequency:	HINTS data were collected in 2003 and 2005. The third administration of the survey was funded in 2006 and will go into the field in 2007.
Data Collection Methodology:	Telephone data collection was conducted using computer-assisted telephone interviewing (CATI) by trained telephone interviewers over a period of 25 weeks. In 2005, some respondents were offered a web-based option for completing the survey using the Internet.
Participation:	Optional, with incentives

Health Information National Trends Survey (HINTS)
(continued)

Response Rate:	The 2005 overall response rate is reported as 20.83 percent.
Sampling Methodology:	The sample design for both the 2003 and 2005 HINTS was a list-assisted, random digit dialing (RDD) sample of all telephone exchanges in the United States. This approach resulted in a nationally representative sample of households. During the household screener, one adult was sampled within each household and recruited for the extended interview. The respondent selection process was the same for the 2003 and 2005 HINTS with a minor modification for three-person households.
Analysis:	The HINTS data set is weighted. Base weights were assigned to both sampled households and sampled adults within households. Standard errors were computed for HINTS 2005 estimates. The HINTS 2005 Final Report contains detailed information regarding the computation of these weights.
Strengths:	Data are collected on health policy issues including health communication and information. There are multiple years of data available.
Limitations:	There are a very limited number of AI/AN/NA individuals in this study. The response rate for the 2005 HINTS study is very low.
Access Requirements and Use Restrictions:	Data can be downloaded directly from the HINTS website at no cost (after registering).
Contact Information:	Data can be downloaded at http://hints.matthewsgroup.com/register.asp. NCI Division of Cancer Control and Population Sciences 6130 Executive Boulevard, Suite 6134 Bethesda, Maryland 20892 http://hints.cancer.gov http://cancercontrol.cancer.gov/hints/

Integrated Postsecondary Education Data System (IPEDS)

Sponsor:	U.S. Department of Education (DoE)/National Center for Educational Statistics (NCES)
Description:	The Integrated Postsecondary Education Data System (IPEDS), established as the core postsecondary education data collection program for NCES, is a system of surveys designed to collect data from all primary providers of postsecondary education. For IPEDS, a postsecondary institution is defined as an organization open to the public that has as its primary mission the provision of postsecondary education (defined as formal instructional programs with a curriculum designed primarily for students who are beyond the compulsory age for high school). The IPEDS system is built around a series of interrelated surveys designed to collect institution-level data in such areas as enrollment, program completion, faculty, staff, and finances. IPEDS surveys postsecondary institutions, including universities and colleges, as well as institutions offering technical and vocational education beyond the high school level. All institutions that participate in any federal student financial assistance program authorized by Title IV must submit data to IPEDS.
Relevant Policy Issues:	Educational Attainment.
Data Type(s):	Program reporting data
Unit of Analysis:	Educational institution
Identification of AI/AN/NA:	Institutions report data on American Indians/Alaska Natives (AI/AN) students in three areas: enrollment, completers by program studied, and graduation rates.
AI/AN/NA Population in Data Set:	Enrollment at Title IV institutions in the U.S., Fall 2004: Total: 17,710,798 AI/AN population = 170,919 Using IPEDS, it is also possible to disaggregate the data by tribal colleges. Thirty-two tribal colleges submitted data to IPEDS in 2004. Enrollment at tribal colleges in the U.S., Fall 2004: Total enrollment at tribal colleges: 17,599 AI/AN enrollment at tribal colleges = 14,067
Geographic Scope:	The geographic scope of IPEDS is national. Institutions can be grouped by state for state-level analysis. Additional information on the state of residence of first-time freshmen is collected in even years.
Date or Frequency:	IPEDS surveys are collected on an annual basis. IPEDS began in 1986, replacing the Higher Education General Education Information Survey (HEGIS) which began in 1966.

Integrated Postsecondary Education Data System (IPEDS)
(continued)

Data Collection Methodology: IPEDS uses a web-based data collection system. Each institution designates a person or several persons who are responsible for ensuring that survey data submitted for the institution are correct. Non-response follow-up is conducted with CEOs, coordinators, and data entry staff via mail, e-mail, and telephone.

Participation: Participation is mandatory for Title IV institutions. Institutions that do not participate in Title IV programs may participate in the IPEDS data collection on a voluntary basis.

Response Rate: As data collection is mandatory for Title IV institutions, response rates are very high. Response rates calculated by type of institution for all IPEDS survey components are above 95 percent. Response rates for non-Title IV institutions are not provided.

Authorization: The completion of the surveys by all institutions that participate in or are applicants for participation in any federal financial assistance program authorized by Title IV of the Higher Education Act of 1965, as amended, is mandated by 20 U.S.C. 1094, Section 487(a)(17).

Strengths: The data sets contain a large number of AI/AN respondents. It is possible to disaggregate the data to examine counts by tribal colleges. Data are collected on a key policy issue, education. There are multiple years of data available.

Limitations: Data are only available at the institutional level.

Access Requirements and Use Restrictions: Data are available on the Internet at no charge.

Contact Information: IPEDS data are available online. The online table generator for IPEDS is located at http://nces.ed.gov/dasol/tables/. Additionally, researchers can download IPEDS data to their PC using the IPEDS Peer Analysis and Data Cutting Tool at http://nces.ed.gov/ipedspas/.

National Center for Education Statistics
Institute of Education Sciences
1990 K Street, NW
8th & 9th Floors
Washington, DC 20006, USA
Telephone: (202) 502-7300

Medicaid Analytic Extract (MAX)

Sponsor: U.S. Department of Health and Human Services (DHHS)/Centers for Medicare and Medicaid Services (CMS)

Description: The Medicaid Analytic Extract (MAX) files (formerly State Medicaid Research Files) are comprised of the Personal Summary, Inpatient, Long-Term Care, Drug, and Other Therapy data sets, and contain eligibility and utilization records. The Personal Summary File contains one record for every individual enrolled in Medicaid for at least one day during the year. These files include demographic data (e.g., date of birth, gender, race); basis of eligibility; maintenance assistance status; monthly enrollment status; utilization summary; complete inpatient stay records; claims for long term care services provided by Nursing Facilities, Skilled Nursing Facilities (SNFs), Intermediate Care Facilities (ICFs), and independent psychiatric facilities; drug claims; and claim records for all non-institutional Medicaid services, including physician services, lab/X-ray, clinic services and premium payments.

Relevant Policy Issues: Measurement of Health Status, Disease-specific Measurements, Health Disparities, Factors Contributing to Measured Health Disparities, Measures of Well-being for Children, and Measures of Well-being for Elders.

Data Type(s): Program reporting data

Unit of Analysis: Individual

Identification of AI/AN/NA: Racial/ethnic categories available:

- White (was “White, not of Hispanic origin” through September 1998)
- Black or African American (was “Black, not of Hispanic origin” through September 1998)
- American Indian or Alaska Native (AI/AN)
- Asian (was “Asian or Pacific Islander” through September 1998)
- Hispanic or Latino – No Race information available (was “Hispanic" [without race annotation] through October 1998)
- Native Hawaiian or other Pacific Islander (NH/PI) (new code beginning October 1998)
- Hispanic or Latino and one or more Races (new code beginning October 1998)
- More than one Race (new code beginning October 1998)
- Unknown

AI/AN/NA Population in Data Set: Unweighted population in data source from 2003 files:
TOTAL: 55,157,775
AI/AN: 806,211
NH/PI: 508,106

Medicaid Analytic Extract (MAX)
(continued)

Geographic Scope:	The geographic scope of the MAX files is national. State, county, and zipcode also are available for analyses. Files for the District of Columbia are available beginning with the MAX 1999.
Date or Frequency:	Data are submitted quarterly, with data files made available on an annual basis from 1992 to the present.
Data Collection Methodology:	States submit data via the Medicaid Statistical Information System (MSIS), meeting standardized specifications. MSIS data are cleaned and reconciled and become the MAX files.
Participation:	Mandatory
Response Rate:	As the submission of these data are mandatory, it is assumed that the response is a near census of all states. This assumption is only valid for reporting of eligibles and Fee for Service utilization. Reporting of encounter data for services provided under a capitated managed care plan was mandated beginning with FY 1999. The data are still viewed as largely incomplete for utilization.
Authorization:	The Balanced Budget Act of 1997 mandated that all states submit Medicaid-paid claims data to CMS. Prior to this, states submitted data on a voluntary basis.
Strengths:	These data files contain a very large number of AI/AN/NA respondents. Data contain indicators concerning key policy issues including health, child well-being, and elder well-being. There are multiple years of data available. Because these data represent a census of Medicaid data, response rates and sample sizes are not an issue in using the data. States submit data on standardized forms and all data are available in electronic form making the data relatively easy to access and use.
Limitations:	The data files are quite large and cumbersome to use. Potential users must also be extremely familiar with the data documentation to ensure that they are interpreting results obtained from this data set correctly. For example, the variation in procedures and practices across states (e.g., one state tracks or defines a certain service differently) means that there are numerous exceptions or variations on the data standardization requirements. The extensive documentation tracking these data anomalies that should be examined by potential users of these data.
Access Requirements and Use Restrictions:	Researchers must submit a proposal and comply with multiple criteria of the data use agreement. Note that only approved academic research projects and certain government agencies are entitled to a data use agreement to obtain MAX data. Detailed information can be found at: http://www.resdac.umn.edu/Medicaid/requesting_data.asp. Cost of the data set is dependent on the number of states, years, and file types requested.

Medicaid Analytic Extract (MAX)
(continued)

Contact Information:	Contact Information: E-Mail: resdac@umn.edu Phone: (888) 9-ResDAC or (888) 973-7322 Data documentation: http://www.resdac.umn.edu/Medicaid/data_documentation.asp. http://www.cms.hhs.gov/MedicaidDataSourcesGenInfo/07_MAXGeneralInformation.asp

Medical Expenditure Panel Survey (MEPS)

Sponsor: U.S. Department of Health and Human Services (DHHS)/Agency for Health Care Research and Quality (AHRQ) and National Center for Health Statistics (NCHS)

Description: The Medical Expenditure Panel Survey (MEPS) is conducted to provide nationally representative estimates of health care use, health care expenditures, sources of payment, health insurance coverage and health status for the U.S. civilian noninstitutionalized population. MEPS data can be used to estimate the impact of changes in sources of payment and insurance coverage on different economic groups or special populations of interest, such as the poor, elderly, families, veterans, the uninsured, and racial and ethnic minorities.

MEPS is comprised of three component surveys: the Household Component (HC), the Medical Provider Component (MPC), and the Insurance Component (IC). The HC is the core survey, and it forms the basis of the MPC sample. Together these surveys yield comprehensive data that provide national estimates of the level and distribution of health care use and expenditures, support health services research, and can be used to assess health care policy implications.

Relevant Policy Issues: Measurement of Health Status, Disease-specific Measurements, Health Disparities, Factors Contributing to Measured Health Disparities, Measures of Well-being for Families/Households, Factors Contributing to Well-being Disparities of Families, Measures of Well-being for Children, Measures of Well-being for Elders, and Factors Contributing to Well-being Disparities of Elders.

Data Type(s): Survey

Unit of Analysis: Data can be analyzed at the person level, the event level, the family level, or the health insurance unit level.

Identification of AI/AN/NA: Race categories include (2002 and beyond):

- White alone
- Black alone
- American Indian or Alaska Native (AI/AN) alone
- Asian alone
- Native Hawaiian or Pacific Islander (NH/PI) alone
- Multiple Race

AI/AN/NA Population in Data Set: Based on 2004 MEPS data, out of 34,403 total records, 293 are coded AI/AN alone and 150 are coded NH/PI alone. Multiple years of data can be pooled to increase the sample size. Aggregation requires at least 100 unweighted cases to support national estimates.

Geographic Scope: The geographic scope of the MEPS is national. The MEPS is designed to support national and regional estimates. Due to small sample sizes, state estimates of the AI/AN/NA population are not possible.

Medical Expenditure Panel Survey (MEPS)
(continued)

Date or Frequency: MEPS was initiated in 1996 and is a continuous ongoing survey. MEPS predecessor surveys were conducted in 1987 and 1977.

Data Collection Methodology: The MEPS HC uses an overlapping panel design in which data are collected through a series of five rounds of interviews over a 2-1/2 year period using computer-assisted personal interviewing technology to collect information on all household members. This series of data collection rounds is launched each subsequent year on a new sample of households to provide overlapping panels of survey data that, when combined, will provide continuous and current estimates of health care expenditures. In 2000, an annual fielding of a self-administered questionnaire (SAQ) was introduced into the process as well.

For the MPC, a sample of medical providers are contacted by telephone to provide information that the household respondents cannot accurately provide (after obtaining permission from the HC respondents). The IC is an annual panel survey that collects data on health insurance plans obtained through employers, unions, and other sources of private health insurance. Data are collected by the Census Bureau from the sampled organizations through a prescreening telephone interview, a mailed questionnaire, and a telephone follow-up for non-respondents.

Participation: Optional, with incentives ($25.00) for each of the five rounds of interviews completed and $5.00 for each self-administered or child questionnaire completed.

Response Rate: The full-year HC response rate has generally ranged from about 65 to 71 percent. Conditional response rates for Rounds 2-5 are always over 90 percent.

Sampling Methodology: The MEPS-HC uses the National Health Interview Survey (NHIS) conducted by the National Center for Health Statistics as its sampling frame. The NHIS sampling frame provides a nationally representative sample of the U.S. civilian non-institutionalized population and reflects an over-sampling of Blacks and Hispanics. In certain years MEPS over samples additional policy relevant sub-groups. This design allows linkage back to the previous year's NHIS for analytic purposes.

Analysis: As with most surveys, participating individuals represent only a fraction of the overall population the survey is intended to reflect. In order to calculate estimates representing the overall population, responses from surveyed individuals must be weighted by the proportion of the population they represent. In addition, adjustments must be made to account for non-response. Each MEPS file contains appropriate weight variables that can be applied to the data to generate national estimates of the civilian non-institutionalized population. A detailed description of the weighting process and how weights are applied to estimates can be found in the weighting and estimation section of the online workbook under the workshop and events section of the MEPS web site.

Because MEPS relies on a complex sampling design rather than simple random sampling, it is important to take into account reductions in the accuracy of

calculated estimates (size of standard errors). Statistical software programs such as SAS (version 8.2 or higher), SUDAAN, STATA, and SPSS (version 12.0 or higher), are available to accommodate the complex design and calculate robust standard errors. A full description of how to compute standard errors for MEPS can be found on the MEPS web site (http://www.meps.ahrq.gov/FactSheets/FS_StandardErrors.HTM).

Strengths: MEPS facilitates research on relationships between individual characteristics and health care utilization. It provides national and regional estimates of health care use and expenditures. It contains two full years' worth of data for each panel. Each panel also can be linked to the previous year's NHIS for a third point in time. This design facilitates analysis of change over time. MEPS data can be used to estimate the impact of changes in sources of payment and insurance coverage on different populations of interest such as the AI/AN/NA population, or to evaluate the impact of an intervention or treatment on health status over time. It is the only source of actual sources of payments and amounts paid including out-of-pocket expenditures for health care visits.

Limitations: The MEPS was not designed to produce state-level estimates. While aggregate estimates for a selected number of large states may be possible, small sample sizes preclude making such estimates for the AI/AN/NA populations. Even after pooling several years of MEPS data, sample size limitations and confidentiality restrictions make MEPS data unsuitable for certain types of analysis. For example, the MEPS data do not support research on rare conditions. Moreover, information on conditions is household reported and not verified by clinical records. All MEPS data are reported by one designated household respondent. Reporting detailed information on other household members can sometimes be problematic.

Access Requirements and Use Restrictions: MEPS HC data releases, including documentation and codebooks, are available free to the public on the Internet (via the MEPS web site). MEPS IC data are published in tabular format on a yearly schedule. MEPS data (HC and IC data) are also available via MEPSnet, an on-line, interactive, statistical tool developed to give users the ability to analyze MEPS data in real-time. Access to the MEPS IC full data set is only available in a Census Bureau Research Data Center.

Many of the MEPS databases include considerably more data than can be made available to the general public because of the constraints of confidentiality guidelines. In order to facilitate the use of such data, while maintaining the confidentiality, AHRQ developed a Data Center (a physical space at AHRQ in Rockville, Maryland) where researchers with approved projects can access data files not available for public dissemination. See the MEPS web site for details.

Medical Expenditure Panel Survey (MEPS)
(continued)

Contact Information:

The MEPS website is http://www.meps.ahrq.gov/.

By e-mail: mepspd@ahrq.gov
By phone/fax: MEPS Information Coordinator (301) 427-1656
CFACT General Information: (301) 427-1406
CFACT Fax: (301) 427-1276

By mail: Project Director
Center for Financing, Access and Cost Trends
Medical Expenditure Panel Survey
Agency for Healthcare Research and Quality
540 Gaither Road
Rockville, MD 20850
Telephone: (301) 427-1656
E-mail: mepspd@ahrq.gov

Reports of Interest

For copies of data products and reports, see the MEPS web site. Selected MEPS data products are available from:
AHRQ Publications Clearinghouse
P.O. Box 8547
Silver Spring, MD 20907
(1-800) 358-9295
(703) 437-2078 outside the U.S.
TDD for the hearing impaired, toll free: (888) 586-6340

Medicare Denominator File

Sponsor: U.S. Department of Health and Human Services (DHHS)/Centers for Medicare and Medicaid Services (CMS)

Description: The Medicare Denominator File contains demographic and enrollment information about each beneficiary enrolled in Medicare during a calendar year. The information in the Denominator File is finalized in March of the following calendar year. Some of the information contained in the file includes the beneficiary unique identifier, state and county codes, zipcode, date of birth, date of death, sex, race, age, monthly entitlement indicators (Part A/B/Both), reasons for entitlement, state buy-in indicators, and monthly managed care indicators (yes/no). The Denominator File is used to determine beneficiary demographic characteristics, entitlement, and beneficiary participation in Medicare Managed Care Organizations.

Relevant Policy Issues: Factors Contributing to Measured Health Disparities and Measures of Well-being for Elders.

Data Type(s): Program enrollment, eligibility, and demographic data

Unit of Analysis: Individual

Identification of AI/AN/NA: Race is provided by the Social Security Administration and is found within the Denominator file. CMS receives a quarterly update file from Indian Health Services on individuals deemed by IHS to be Native American and uses it to update the race code found in the Denominator file. The information from the IHS will override any other code which may have been on file.

Race/ethnic categories included in the data are:

- Unknown
- White
- Black
- Other
- Asian/Pacific Islander
- Hispanic
- North American Native

AI/AN/NA Population in Data Set: Population size is not available in documentation, however, this file is a census of the Medicare population and should have more than sufficient sample size for research purposes. The Native American population from the administrative records in 2002, for example, was over 141,000.

Geographic Scope: The geographic scope of the file is national. State, county, and zip code are available in the file.

Medicare Denominator File
(continued)

Date or Frequency:	Data are collected on an ongoing basis with the files constructed on an annual basis.
Data Collection Methodology:	The Medicare Denominator Files are populated using the Medicare enrollment database, an administrative data source completed when individuals become eligible for and enroll in the Medicare program. An update in March of each year is included for the previous year's data file (e.g., March 1995 update is used to finalize the 1994 denominator file).
Participation:	Mandatory
Strengths:	The Medicare Denominator File can be combined with other data sources (e.g., Standard Analytic File/Claims data) using the beneficiary's unique identifier, and therefore can be a very powerful analytic data source. On their own, these files offer a census of Medicare enrollment, type of coverage and basic demographics. They contain a large number of AI/AN/NA respondents and there are multiple years of data available.
Limitations:	The Pacific Islander population is not separated from the Asian racial/ethnic category. The AI/AN/NA populations are also not divided into more refined categories. Given the size of the Medicare population, cell size concerns should not be an issue, yet only these broader race/ethnic categories are available. This is due in large part to how the data are collected.
Other:	In addition to the data sets profiled in this catalog, a full listing of all available CMS data sets is presented at this website: http://www.cms.hhs.gov/FilesForOrderGenInfo/
Access Requirements and Use Restrictions:	The data request process involves multiple components including a proposal of how the data will be used, evidence of funding for the research, data use agreement, Institutional Review Board clearance/waiver, and some administrative forms. There is a cost associated with use of this data set. Researchers will need to contact ResDAC in order to receive pricing information. The steps and forms are described in detail at the following website: http://www.resdac.umn.edu/medicare/requesting_data.asp.
Contact Information:	The University of Minnesota, CMS contractor for managing and supporting CMS data, can be contacted via the Internet at: https://resdac.oit.umn.edu/ Via e-mail: resdac@umn.edu Via telephone: (888) 973-7322 Denominator File Record Layout http://www.resdac.umn.edu/ddde/dd_de.asp

Medicare Utilization – Standard Analytic Files (SAFs)

Sponsor: U.S. Department of Health and Human Services (DHHS)/Centers for Medicare and Medicaid Services (CMS)

Description: The Standard Analytical Files (SAFs) comprise seven data sets containing detailed claims information about health care services rendered to Medicare beneficiaries in fee-for-service Medicare. SAFs are available for each institutional (inpatient, outpatient, skilled nursing facility, hospice, or home health agency) and non-institutional (physician and durable medical equipment providers) claim types. Data are organized at the claim level and include basic beneficiary demographic information, date of service, diagnosis and procedure code, provider number, and reimbursement amount.

Relevant Policy Issues: Disease-specific Measurements, Health Disparities, and Factors Contributing to Measured Health Disparities.

Data Type(s): Program claims data

Unit of Analysis: Final action claims, which are associated with an individual.

Identification of AI/AN/NA: Race is provided by the Social Security Administration and is found within the Denominator file. CMS receives a quarterly update file from Indian Health Services on individuals deemed by IHS to be Native American and uses it to update the race code found in the Denominator file. The information from the IHS will override any other code which may have been on file.

The following categories are used in the SAFs:

- White
- Black
- Other
- Asian
- Hispanic
- North American Native
- Unknown

AI/AN/NA Population in Data Set: Population size is not available in documentation, however, the data reflect all Medicare beneficiaries and therefore should have more than adequate sample size to address research needs.

Geographic Scope: The geographic scope of the SAFs is national. Zip code is the lowest level of geographic detail available in the file.

Date or Frequency: Data are submitted continually from the Fiscal Intermediary (FI) and Carriers to CMS, but SAFs are produced by calendar year, available from 1991 through the present, with 2004 data the most recently available.

Medicare Utilization – Standard Analytic Files (SAFs)
(continued)

Data Collection Methodology: Providers submit claims to the FI or Carrier for processing and payment. The FI or Carrier forwards all claims to CMS. CMS creates the SAFs six months following the end of the calendar year.

Participation: Mandatory

Strengths: This data source offers a near-census (does not include the small number of beneficiaries enrolled in managed care organizations) of utilization, expenditure and diagnosis data for Medicare beneficiaries.

CMS updates race information based on a quarterly update file from Indian Health Services identifying individuals deemed by IHS to be Native American.

Limitations: Only very broad racial categories are available (White, Black, Hispanic, Asian, Other, North American Native). Claims data are very difficult to work with and also consist of very large files, often millions of records large, and therefore are not a suitable resource for the average researcher with limited programming skills.

Other: In addition to the data sets profiled in this catalog, a full listing of all available CMS data sets is presented at this website:
http://www.cms.hhs.gov/FilesForOrderGenInfo/

Access Requirements and Use Restrictions: The data request process involves multiple components including a proposal of how the data will be used, evidence of funding for the research, data use agreement, Institutional Review Board clearance/waiver, and some administrative forms. There is a cost associated with use of this data set. Researchers will need to contact ResDAC in order to receive pricing information. The steps and forms are described in detail at the following website:
http://www.resdac.umn.edu/medicare/requesting_data.asp.

Contact Information: The University of Minnesota, CMS contractor for managing and supporting CMS data, can be contacted via the Internet at: https://resdac.oit.umn.edu/
Via e-mail: resdac@umn.edu
Via telephone: (888) 973-7322

Please see this website for a further description of the SAFs:
http://www.resdac.umn.edu/Medicare/data_file_descriptions.asp#rif

Data Dictionaries are available at:
http://www.resdac.umn.edu/ddvh/index.asp#RIF

National Aging Program Information Systems (NAPIS) State Program Reports

Sponsor: U.S. Department of Health and Human Services (DHHS)/Administration on Aging (AoA)

Description: The National Aging Program Information Systems (NAPIS) State Program Reports are completed by the states to comply with AoA reporting requirements for submission of annual performance reports. Three principal types of data are included in the NAPIS design: (1) performance data on programs and services funded by the Older Americans Act (OAA); (2) demographic/descriptive data on the elderly population obtained from the U.S. Census Bureau and other sources; and (3) descriptive data on the infrastructure of home- and community-based services in place to assist older persons, based on AoA studies and related reviews.

Relevant Policy Issues: Factors Contributing to Well-being Disparities of Elders.

Data Type(s): Program reporting data

Unit of Analysis: Primarily aggregated state-level data. Under Title VI of the Older Americans Act, grant awards are made directly to tribal and native organizations representing older American Indians, Alaska Natives, and Native Hawaiians. Title VI data will be available by individual grantee.

Identification of AI/AN/NA: On the reporting form entitled *General Characteristics of Elderly Clients Receiving Registered Services and Those Receiving Cluster 2 Registered Services*, respondents are given the following options for reporting clients by race or ethnicity:

- White (Alone) – Non-Hispanic
- White (Alone) - Hispanic
- American Indian or Alaska Native (AI/AN) (Alone)
- Asian (Alone)
- Black or African American (Alone)
- Native Hawaiian or other Pacific Islander (NH/PI) (Alone)
- Persons Reporting Some Other Race
- Persons Reporting Two or More Races
- Race Missing

AI/AN/NA Population in Data Set: The total number of persons served under Title III of the OAA for FY 2004 is 8,651,974. Of these, the total number of AI/AN individuals in the FY 2004 data is 56,606. NH/PI is combined with Asians in all published reports.

Geographic Scope: The geographic scope of the study is national. The identified geographic areas are states.

Date or Frequency: States submit performance reports annually.

National Aging Program Information Systems (NAPIS) State Program Reports (continued)

Data Collection Methodology:	States submit data to NAPIS through an online reporting application.
Participation:	Mandatory
Authorization:	The 1992 reauthorization of the Older Americans Act (OAA) directed the AoA to improve performance reporting on programs and services funded by the OAA. This impetus caused AoA to reconsider reporting requirements for all Titles of the Act under the direction of AoA. The concept of the National Aging Program Information System (NAPIS) arose from this review and the related review of AoA's internal information needs.
Strengths:	State information contains a large number of AI/AN/NA respondents. Data are collected on the key policy issue of elder well-being. There are multiple years of data available.
Limitations:	Most data are only available at the state level.
Access Requirements and Use Restrictions:	Data set is available to public at no cost.
Contact Information:	The general website for the NAPIS program is: http://www.aoa.gov/prof/agingnet/NAPIS/napis.asp Questions regarding NAPIS should be addressed to: Saadia Greenberg email: saadia.greenberg@aoa.gov or Steve Cordasco email: steve.cordasco@aoa.gov To obtain the NAPIS data, contact: Robert Hornyak email: robert.hornyak@aoa.hhs.gov

National Ambulatory Medical Care Survey (NAMCS)

Sponsor: U.S. Department of Health and Human Services (DHHS)/Centers for Disease Control and Prevention (CDC)/National Center for Health Statistics (NCHS)

Description: The National Ambulatory Medical Care Survey (NAMCS) collects information about the provision and use of ambulatory medical care services in the United States. Non-federally employed office-based physicians complete a one-page questionnaire for each patient visit sampled during a one-week reporting period. Collected data include physician characteristics (obtained during a survey induction interview), patient demographic characteristics (age, sex, race, ethnicity), and visit characteristics (patients' symptoms, complaints or other reasons for the visit, physicians' diagnoses, diagnostic and therapeutic services ordered or provided at the visit including medications, expected sources of payment, visit disposition, time spent with physician, etc.). Participating physicians must be primarily involved in office-based direct patient care, with anesthesiologists, pathologists and radiologists excluded.

Relevant Policy Issues: Measurement of Health Status, Key Health Disparities, and Factors Contributing to Measured Health Disparities.

Data Type(s): Survey

Unit of Analysis: Patient visit

Identification of AI/AN/NA: Physicians are asked to report one or more races (up to 5) for each sampled visit. The public use data file includes five single race categories and an aggregated category for visits with more than one race checked.

- White
- Black/African American
- Asian
- Native Hawaiian/Other Pacific Islander (NH/PI)
- American Indian/Alaska Native (AI/AN)
- More than one race reported

AI/AN/NA Population in Data Set: Total number of records in 2004 data set: 25,286
AI/AN: 93
NH/PI: 70

Geographic Scope: The geographic scope of the study is national. Analysis is possible for the following regions: Northeast, Midwest, South, West.

Date or Frequency: Data are available annually from 1973 to 1981, in 1985, and annually since 1989.

National Ambulatory Medical Care Survey (NAMCS)
(continued)

Aggregation:	Each year of data, on average, has 50-90 visits by persons reported as AI/AN only and 60-100 visits by persons reported as NH/PI only. According to the NCHS, researchers frequently combine years of data for analysis in order to achieve reliable estimates. Researchers considering aggregation should take special note of changes in sample design variables across the years, as these will affect variance estimation. They should also be particularly aware of any possible clustering by race that may affect sample estimates. The format and content of the survey questionnaires has also changed across the years. Data must be weighted to produce national estimates, and researchers may wish to seek guidance about the use of weights with aggregated files.
Data Collection Methodology:	The U.S. Census Bureau acts as the data collection agent for the NAMCS. The physician, or his/her staff, is trained by Census field representatives to sample patients and to complete the 1-page reporting form for each sampled visit.
Participation:	Optional, without incentives
Response Rate:	In the 2004 survey, the response rate for participating physicians was 64.7 percent.
Sampling Methodology:	NAMCS utilizes a multistage probability sample design where geographic primary sampling units (PSUs) are selected in the first stage, physician practices within PSUs in the second stage (using the American Medical Association and American Osteopathic Association directories as the sampling frame), and a random sample of patient visits to selected physicians in the third stage.
Analysis:	The weighting procedure produces essentially unbiased national estimates and has basically four components: 1) inflation by reciprocals of the probabilities of selection, 2) adjustment for nonresponse, 3) a ratio adjustment to fixed totals, and 4) weight smoothing. Patient visit weights are provided in the data set to produce accurate national estimates.
Strengths:	Data are collected on key policy issues pertaining to health. There are multiple years of data available.
Limitations:	In the 2004 survey, the item nonresponse rate is low overall; however this rate for the variables measuring ethnicity and race is 20.9 percent and 19.2 percent, respectively. Race and ethnicity are imputed from records with similar characteristics based on physician specialty, geographic region (in the case of ethnicity, state is used rather than region), and primary diagnosis. There are also few visits by patients categorized as AI/AN or NH/PI. Finally, NAMCS does not include physicians from Indian Health Service (IHS) in the sample frame.
Access Requirements and Use Restrictions:	Data are available to the public at no cost. Restricted files which contain additional variables and non-masked data can be accessed by applying to the NCHS Research Data Center and paying a fee.

National Ambulatory Medical Care Survey (NAMCS)
(continued)

Contact Information:

Main website for NAMCS
http://www.cdc.gov/nchs/namcs.htm

Data Download:
http://www.cdc.gov/nchs/about/major/ahcd/ahcd1.htm#Micro-data

Contact Information:
National Center for Health Statistics
Ambulatory Care Statistics Branch
3311 Toledo Road, Rm. 3409
Hyattsville, MD 20782
(301) 458-4600

National Assessment of Adult Literacy (NAAL)

Sponsor: U.S. Department of Education (DoE)/National Center for Education Statistics (NCES)

Description: The National Assessment of Adult Literacy (NAAL) is a nationally representative assessment of English literacy among American adults age 16 and older. Results from the study cover the status and progress of literacy in the nation, the literacy skill level of American adults (including the least-literate adults), various factors associated with literacy, and the application of literacy skills to health-related materials. NAAL also provides the results of state-level assessments of six participating states and a national study on literacy among the state and federal prison population (local jails and other types of institutions are not included). Additionally, one important goal of the 2003 NAAL was to provide trend data in adult literacy performance since 1992. At this time, the 2003 NAAL full data set and accompanying technical documentation are not ready for public release.

Relevant Policy Issues: Educational Attainment.

Data Type(s): Survey

Unit of Analysis: Individual

Identification of AI/AN/NA: During the in-person interview, respondents are asked "Which of the groups on this card best describes you? Choose one or more." and then given a handcard with the following response options:

- White
- Black or African American
- Asian
- American Indian or Alaska Native (AI/AN)
- Native Hawaiian or other Pacific Islander (NH/PI)

AI/AN/NA Population in Data Set: Of the 19,714 adults who made up the 2003 NAAL sample, 18,541 were from the household sample and 1,173 were from the prison sample. The total number of AI/AN respondents included in the full 2003 NAAL sample is 167. The total number of NH/PI respondents included is 26. An upcoming NAAL report will include separate estimates for AI/ANs.

Geographic Scope: The geographic scope of the study is national. In addition to allowing for national estimates, both the 1992 and 2003 NAAL allow estimates for state-level assessments for those states that choose to participate. In 2003, the following states participated in state-level assessments: Kentucky, Maryland, Massachusetts, Missouri, New York, and Oklahoma.

National Assessment of Adult Literacy (NAAL)
(continued)

Date or Frequency: The NAAL was conducted in 1992 and 2003. In order to provide trend data on adult literacy in the future, NCES plans to conduct additional periodic assessments.

Aggregation: The 1992 and 2003 NAAL used the same sampling and data collection procedures to ensure that comparable populations were assessed in both years. However, there are differences in how race information was collected between the two administrations. In 2003, respondents were not offered an "Other" category to describe their race and they could select one or more racial categories, while in 1992 respondents were limited to choosing one race or ethnicity. Additionally, there is an eleven year gap between the two data collection efforts, which may contribute to the variability in responses. For these reasons caution should be exercised if combining the 1992 and 2003 results for analysis by race/ethnicity.

Data Collection Methodology: NAAL is administered in person. Participants are first administered a set of background questions using computer assisted personal interviewing (CAPI) and a set of basic screening tasks to determine whether they should be given the main NAAL or the Adult Literacy Supplemental Assessment (ALSA). The least-literate participants are given the ALSA. Main NAAL participants read assessment questions from printed booklets and write their answers using a pencil. ALSA participants give oral responses to oral questions, but refer to printed materials to find the answers. At the end of the interview, all participants take the Fluency Addition to NAAL (FAN), which requires them to read lists and passages aloud from printed booklets. Participants' responses to FAN are recorded using CAPI software.

Participation: Optional, with incentives. Respondents were paid $30 to participate.

Response Rate: Cases are considered complete if the respondent completed the background questionnaire and at least one question on each of the three scales of literacy assessment, or if the respondent was unable to answer questions due to language issues or mental disabilities. The overall weighted response rate for the household sample was 60.1 percent. The overall weighted response rate for the prison sample was 87.2 percent.

Sampling Methodology: NAAL uses a multi-stage probability sampling design. The NAAL sample is designed to represent all U.S. adults who live in households and prisons. For the 2003 NAAL, a national sample of the adult household population was combined with samples for the six states that participated in the NAAL state-level assessment. Stage 1 of the sample design involved dividing the U.S. into 1,900 primary sampling units (PSUs) and selecting 160 PSUs total among the national and state samples combined. Stage 2 involved selecting area segments within each selected PSU. Stage 3 was the selection of sampled households, and stage 4 was the selection of individual participants within the households.

National Assessment of Adult Literacy (NAAL)
(continued)

The NAAL prison sample was independently selected using a two-stage design. The first stage involved selecting more than 100 federal and state prisons. The second stage involved selecting individual inmates from the selected prisons.

Strengths: Data are collected on a key policy issue, adult literacy. There are multiple years of data available.

Limitations: There are a small number of AI/AN/NA respondents. Differences in how race was assessed in 1992 and 2003, as well as the length of time between the two data collection efforts, will affect the ability to aggregate the data. Although currently there is a public use data set containing a limited number of variables available on the NAAL website, this limited data set does not identify AI/AN individuals (available race categories in the public use data set include White, Black, Hispanic, and Other.)

The full NAAL public-use and restricted-use data sets are not ready for dissemination at this time. However, once they are released, only the restricted use NAAL data set will identify AI/AN individuals. As there are less than 45 NH/PI cases in the data set, these values will be suppressed according to NAAL rules.

Access Requirements and Use Restrictions: To access the NAAL restricted-use data set, individuals will have to obtain an NCES license. NAAL staff currently will respond to special requests for analysis of NAAL data. Users who want to request that service need to complete a request form and submit it to NAAL. There is no cost for these analyses.

Contact Information:
National Center for Education Statistics
1990 K Street, NW
Room 8087
Washington, DC 20006
http://nces.ed.gov/NAAL/

National Child Abuse and Neglect Data System (NCANDS)

Sponsor: U.S. Department of Health and Human Services (DHHS)/Administration on Children and Families (ACF)/Children's Bureau

Description: The National Child Abuse and Neglect Data System (NCANDS) is a federally sponsored national data collection effort created for the purpose of tracking the volume and nature of child maltreatment reporting each year within the United States. The NCANDS Child File consists of child-specific data of all investigated reports of maltreatment to state child protective service agencies. Child File data are collected annually through the voluntary participation of states. Participating states submit their data after going through a process in which the state's administrative data system is mapped to the NCANDS data structure. Data elements include the demographics of children and their perpetrators, types of maltreatment, investigation or assessment dispositions, risk factors, and services provided as a result of the investigation or assessment.

Relevant Policy Issues: Measures of Well-being for Children, Child Maltreatment Rates.

Data Type(s): Registry

Unit of Analysis: The unit of observation in the Child File includes report-level data for all children who have received a disposition of an investigation or assessment of allegations of maltreatment during the reporting year. Each child on a report gets a separate data record, referred to as a "report-child pair." As a child may be in the data file multiple times, there is a unique identifier assigned to each child.

Identification of AI/AN/NA: Race is coded by the state agency submitting the data to NCANDS. Beginning with the year 2000, the agency was allowed to select more than one race for a child. Each of the five race variables are independent (White, African American, American Indian or Alaska Native (AI/AN), Native Hawaiian or other Pacific Islander (NH/PI), and Asian), so an individual may have more than one race variable coded as true.

The directions for coding a child as American Indian or Alaska Native are: A child having origins in any of the original peoples of North and South America (including Central America) and who maintains tribal affiliation or community attachment.

The directions for coding a child as Native Hawaiian or other Pacific Islander are: A child having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.

If there are a very small number of records at the county level for a particular race, the race information is recoded to unknown.

AI/AN/NA Population in Data Set:

The 2004 data set consists of 3,134,026 total records (report-child pairs) from 44 states and the District of Columbia (DC).

Counts of distinct children by race in the 2004 NCANDS data file:
AI/AN: 46,708
NH/PI: 11,700

Counts of distinct perpetrators by race in the 2004 NCANDS data file:
AI/AN: 6,294
NH/PI: 2,091

The 2003 data set consists of 1,216,626 total records (report-child pairs) from 22 states and DC.

Counts of distinct children by race in the 2003 NCANDS data file:
AI/AN: 22,228
NH/PI: 3,400

Counts of distinct perpetrators by race in the 2003 NCANDS data file:
AI/AN: 3,740
NH/PI: 653

Geographic Scope:

Forty-four states and DC voluntarily submitted data to the NCANDS Child File for 2004. States that did not submit data for 2004 are Alaska, Alabama, Georgia, North Dakota, Oregon, and Wisconsin.

In 2003, only twenty-two states and DC agreed to archive their NCANDS Child File data with the National Data Archive on Child Abuse and Neglect (NDACAN). States that submitted data in 2003 are Arkansas, Delaware, Florida, Kansas, Kentucky, Louisiana, Massachusetts, Maine, Minnesota, Missouri, Montana, North Carolina, Nebraska, Ohio, Oklahoma, Pennsylvania, Rhode Island, Texas, Utah, Vermont, Washington, and Wyoming.

In addition to the state indicator, the child data file also includes the Federal FIPS Code for the county where the report was made for counties with more than 1,000 records in the data file. For all records, the child's county of residence is removed from the data file because of confidentiality concerns.

Date or Frequency:

The most recent available data are for Federal Fiscal Year 2004. NCANDS data have been collected annually since 1990. For 1990 through 2002, annual data sets are for calendar years. Beginning with the 2003 data set, the collection period is Federal Fiscal Year.

Data Collection Methodology:

Reports of child maltreatment are received by state agencies that administer social services. A single report may contain information about multiple children, multiple abuse types (e.g., physical, sexual, neglect), and multiple perpetrators. The agency investigates the report and a decision is made regarding each

report/child/instance. If the abuse is corroborated by evidence, the report/child/instance is coded as "substantiated" (states vary as to what word they use for this concept). If there is not sufficient corroboration, the instance is not coded as substantiated. Both substantiated and unsubstantiated cases are included in NCANDS. For unsubstantiated cases, no information is collected about the perpetrator. For substantiated cases, the gender, race, relationship to child, and other information is collected about the perpetrator or perpetrators.

State participation in the detailed case data collection consists of mapping each requested data element into the Child File record layout, extracting the state data into the Child File record layout, and submitting the case level data to NCANDS.

Participation: Optional without incentives.

Sampling Methodology: The NCANDS Child File represents a census of all child protective services investigations or assessments conducted in the states that participated in the NCANDS.

Authorization: The Child Abuse Prevention and Treatment Act (CAPTA) was amended in 1988 to direct the Secretary of the Department of Health and Human Services (DHHS) to establish a national data collection and analysis program that would make available state child abuse and neglect reporting information (42 U.S.C. 5101 et seq.; 42 U.S.C. 5116 et seq., Public Law 100-294 passed April 25, 1988). DHHS responded by establishing NCANDS as a voluntary, national reporting system.

Strengths: Data sets contain a large number of AI/AN/NA respondents. Data are collected on a key policy issue, child welfare. There are multiple years of data available.

Limitations: As states are not required to submit data to NCANDS, some states do not participate. Coverage has improved from 2003 (22 states and DC submitting) to 2004 (44 states and DC submitting).

When conducting analyses with NCANDS data, it is important to keep in mind that state-to-state variation in child maltreatment laws and information systems may affect the interpretation of the data. Users are encouraged to refer to the state mapping documents included on the data CD for information about how the state's system codes its data.

Access Requirements and Use Restrictions: Restricted usage files of state report-level data are available for researchers from the National Data Archive on Child Abuse and Neglect at www.ndacan.cornell.edu. Researchers who would like to use the data must fulfill eligibility criteria, submit an application for approval to the Archive, and enter into a legally-binding data license that outlines the requirements for appropriate use of the data. Only individuals holding a faculty appointment or research position at an institution of higher education, a research organization, or a government agency are eligible to obtain the Child File. There is no cost for access to these data.

National Child Abuse and Neglect Data System (NCANDS)
(continued)

Contact Information:	National Data Archive on Child Abuse and Neglect Beebe Hall – FLDC Cornell University Ithaca, New York 14853-4401 (607) 255-7799 ndacan@cornell.edu www.ndacan.cornell.edu

National Crime Victimization Survey (NCVS)

Sponsor: U.S. Department of Justice (DoJ)/Bureau of Justice Statistics (BJS)

Description: The National Crime Victimization Surveys (NCVS) series, previously called the National Crime Surveys (NCS), has been collecting data on personal and household victimization through an ongoing survey of a nationally representative sample of residential addresses since 1973. The four primary objectives of the effort include: 1) to develop detailed information about the victims and consequences of crime; 2) to estimate the number and types of crimes not reported to the police; 3) to provide uniform measures of selected types of crimes; and 4) to permit comparisons over time and geographic areas. Basic demographic information such as age, race, gender, and income is also collected to enable analysis of crime by various subpopulations.

Relevant Policy Issues: Rates of Involvement with Justice System, Lifetime Probability of Being a Victim of a Violent Crime, Lifetime Probability of Being a Victim of a Non-violent Crime, Domestic Violence Rates, Child Maltreatment Rates, Factors Contributing to Disparities in Involvement with Justice System and Outcomes.

Data Type(s): Survey

Unit of Analysis: Individuals, Households, Crime Incidents

Identification of AI/AN/NA: Respondents are allowed to select all race categories that apply from the following:

- White
- Black/African American
- American Indian/Alaska Native (AI/AN)
- Native Hawaiian/Other Pacific Islander (NH/PI)
- Asian
- Other – Specify

AI/AN/NA Population in Data Set: In the combined incident-level file for 1992-2004, there are 162,736 incidents of which 1,621 are incidents where the informant for the household in which there had been an incident is AI/AN (alone or as part of a multiple race designation). Native Hawaiians and Other Pacific Islanders cannot be distinguished from Asians in most of the concatenated incident-level file; however, beginning in 2003, a new race categorization was adopted that allows NH/PI persons to be identified from that time forward. In this incident-level file, there are 71 incidents where the informant was NH/PI (alone or as part of a multiple race designation).

AI/AN/NA Subpopulations: It is possible to isolate individuals who identify themselves as American Indian or Alaska Native and indicate that they reside on Indian lands. (Question 12 asks "Are your living quarters located on an American Indian reservation or on Indian lands?")

National Crime Victimization Survey (NCVS)
(continued)

Geographic Scope:	The geographic scope of the study is national. Geographic identifiers include urban or rural; region; and central city of a Metropolitan Statistical Area (MSA), MSA but not in central city, or not MSA. Geographic analysis is possible for all of these. Also, as noted previously, it is also possible to identify individuals reporting that their residence is on Indian lands.
Date or Frequency:	The NCVS is a semiannual study with data available beginning 1973. Respondents are interviewed every 6 months for a total of seven interviews over a 3-year period. After the seventh interview the household leaves the panel and a new household is rotated in to the sample.
Data Collection Methodology:	The first and fifth interviews are face-to-face; the rest are by telephone.
Participation:	Optional, without incentives
Response Rate:	Survey documentation states that the NCS and NCVS have consistently obtained a response rate of about 95 percent.
Sampling Methodology:	The NCVS sample consists of approximately 50,000 sample housing units selected with a stratified, multi-stage cluster design. The first stage consists of selecting a sample of Enumeration Districts (ED's) from designated Primary Sampling Units (counties, groups of counties, or large metropolitan areas). (ED's are established for each decennial Census and are geographic areas ranging in size from a city block to several hundred square miles, and usually encompassing a population of 750 to 1,500 persons.) In the second stage, each selected ED is divided into segments (clusters of about four housing units each), and a sample of segments is selected. The sample of housing units is divided into six rotation groups, and each group is interviewed every six months for a period of three-and-a-half years.
Analysis:	Use of standard statistical tests with these data would not be accurate because these tests assume independence among observations and a simple, random sample design. The NCVS uses a complex, clustered sampling design in which observations are not independent. Survey documentation presents instructions for two methods to calculate variances for NCVS data that avoid these problems, computing generalized variances and using direct variance calculation methods designed for complex survey design.
Strengths:	Data are collected on key policy issues, primarily justice system issues. There are multiple years of data available. Documention is strong and there is a user-friendly on-line analytical tool for the incident-level data.
Limitations:	The full NVS data set is hierarchical. It is not a flat individual-level data set as are most survey data sources. The file is organized into a hierarchical format which corresponds to variations in household composition and in the occurrence of incidents of victimization. Hierarchical data sets have varying record lengths,

and each record is stored sequentially in the data file. Hierarchical storage is a benefit as it greatly reduces the size and space needed to store and process the data; however, stronger programming skills may be required to correctly analyze a file of this type compared to the traditional file.

Other:

Data are available on-line in aggregated form. Data for 1992-2004 are available for downloading. Data for 1992-2003 can be analyzed on-line (incident-level file) or downloaded.

A series of changes has occurred recently with modifications to the questions, placing more emphasis on "hot topics" of crime, which may provide challenges for multi-year comparisons. Some topics included in 2004 were identity theft, credit card theft and multiple crime situations involving personal information.

Access Requirements and Use Restrictions:

The data are available to the public at no cost.

Contact Information:

For data and documentation:
National Archive of Criminal Justice Data
ICPSR
Institute for Social Research
P.O. Box 1248
Ann Arbor, MI 48106
USA
(800) 999-0960
(734) 998-9825
Data can be accessed at
http://webapp.icpsr.umich.edu/cocoon/NACJD-STUDY/04276.xml.

General information:
Bureau of Justice Statistics
810 Seventh Street, NW
Washington, DC 20531
USA
(202) 307-0765
askbjs@usdoj.gov

Reports of Interest:

The National Archive of Criminal Justice Data provides lists of publications from the NVS data. A list of publications for the 1992-2003 data can be found at http://www.icpsr.umich.edu/cgi-bin/CITATIONS/search?study=3995&method=study&path=NACJD.
A list of publications for the 1992-2004 data can be found at http://www.icpsr.umich.edu/cgi/CITATIONS/search?study=4276&method=study&path=NACJD.

National Epidemiologic Survey on Alcohol and Related Conditions (NESARC)

Sponsor: U.S. Department of Health and Human Services (DHHS)/National Institutes of Health/National Institute on Alcohol Abuse and Alcoholism (NIAAA)

Description: The National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) was designed to determine the magnitude of alcohol use disorders and their associated disabilities in the general population and in subgroups of the population and to examine changes over time in alcohol use disorders and their associated disabilities. It is a longitudinal survey with its first wave of interviews fielded in 2001-2002 and second wave in 2004-2005. The NESARC is a representative sample of the non-institutionalized U.S. population 18 years of age and older.

The NESARC collects data on background, alcohol consumption, alcohol abuse and dependence, alcohol treatment utilization, family history of alcoholism, tobacco use and dependence, medicine use, drug abuse and dependence, drug treatment utilization, family history of drug abuse, major depression, family history of major depression, dysthymia, mania and hypomania, panic disorder and agoraphobia, social phobia, specific phobia, anxiety disorder, personality disorders, antisocial personality disorder, family history of antisocial personality disorder, pathological gambling, medical conditions, and victimization. Public use data are currently available for the first wave of data collection.

Relevant Policy Issues: Key Health Disparities, Factors Contributing to Measured Health Disparities.

Data Type(s): Survey

Unit of Analysis: Individual

Identification of AI/AN/NA: The interviewer presents a flashcard with racial categories listed and says, "Please select 1 or more categories to describe your race."

The respondent chooses one or more from the following categories:

- American Indian or Alaska Native (AI/AN)
- Asian
- Black or African American
- Native Hawaiian or other Pacific Islander (NH/PI)
- White

AI/AN/NA Population in Data Set: In the first wave of data collection (2001-2002), there were 43,093 respondents. The unweighted frequencies for selected racial categories were as follows:

AI/AN: 1,304
NH/PI: 363
Note that respondents could select more than one category.

National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) (continued)

	For analytical purposes, the Census Bureau imputed race for individuals for whom that information was missing. The Bureau also used an algorithm to code a single race category for those individuals who identified themselves as multi-racial. These computations and imputations resulted in a constructed variable where the unweighted count of American Indians or Alaska Natives (not Hispanic) was 701. Native Hawaiians or other Pacific Islanders are combined with Asians as one category of the constructed variable.
Geographic Scope:	The geographic scope of the survey is national. Geographic indicators are available for Census region, Census division, central city vs. not central city, and state. These geographic factors can be incorporated as variables in analyses. However, valid analyses cannot be conducted within these geographic areas because the NESARC was designed to be a representative sample of the U.S.
Date or Frequency:	This is a longitudinal study. The first wave of data collection occurred in 2001-2002 and the second wave occurred in 2004-2005.
Data Collection Methodology:	Data are collected through computer-assisted personal interviews (CAPI).
Participation:	Optional, with incentives. Participants who completed the survey were given $80.
Response Rate:	The overall survey response rate for the NESARC was 81 percent.
Sampling Methodology:	The NESARC used a three-stage sampling design. The sampling frame for the NESARC sample of housing units is the Census 2000/2001 Supplementary Survey (C2SS), a national survey of 78,300 households per month. A group quarters frame was also used. Stage 1 was primary sampling unit (PSU) selection using the C2SS PSUs. Stage 2 was household selection from the sampled PSUs. In Stage 3, one sample person was selected from each household.
Strengths:	Data are collected on key policy issues including alcohol and substance use as well as health. When available, the longitudinal data will provide significant opportunities for examining patterns of alcohol use and disability by individuals.
Limitations:	Change in sample size for the AI/AN/NA population related to attrition from the study across the two waves of data collection is unknown at this time.
Other:	The second wave of data collected in 2004 and 2005 are expected to be ready for use in the summer of 2007.
Access Requirements and Use Restrictions:	First wave data are available to public at no cost.

National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) (continued)

Contact Information:	Public use data file and documentation are available at http://niaaa.census.gov/. Ms. Nekisha Lakins CSR, Incorporated Phone: (703) 741-7157 Fax: (703) 312-5230 Email: nlakins@csrincorporated.com
Reports of Interest:	The link below provides a list of more than 30 publications using the NESARC data: http://niaaa.census.gov/publications.html

National Health Interview Survey (NHIS)

Sponsor: U.S. Department of Health and Human Services (DHHS)/Centers for Disease Control and Prevention (CDC)/National Center for Health Statistics (NCHS)

Description: The National Health Interview Survey is a household survey that serves as the primary source of information on the health of the U.S., non-institutionalized, civilian population. Though it has undergone various changes since its inception in 1957, the NHIS has remained largely unchanged since 1997 and comprises the Basic Core Questionnaire (Family, Sample Adult and Sample Child surveys) as well as topic-focused supplements. The Basic Core collects information on household composition and sociodemographic characteristics, tracking information, information for matches to administrative data bases, and basic indicators of health status, activity limitations, injuries, health insurance coverage, and access to and utilization of health care services. Supplement topics are often selected to coincide with monitoring areas of public health interest and to help in the monitoring of national health goals (e.g., Healthy People 2010). Examples of supplement topics include cancer, diabetes, mental health and complementary and alternative medicine.

Relevant Policy Issues: Measurement of Health Status, Disease-specific Measurements, Factors Contributing to Measured Health Disparities, Measures of Well-being for Families/Households, Factors Contributing to Well-being Disparities of Families, Measures of Well-being for Children, Measures of Well-being for Elders, and Factors Contributing to Well-being Disparities of Elders.

Data Type(s): Survey

Unit of Analysis: Individual, Family, and Household

Identification of AI/AN/NA: The race categories available in the public use data set are:

- White only
- Black/African American only
- American Indian/Alaska Native only (AI/AN)
- Asian only
- Race group not releasable
- Multiple race

AI/AN/NA Population in Data Set: AI/AN only: 670

Geographic Scope: The geographic scope of the NHIS is national. Geographic analysis also is possible at the regional level.

National Health Interview Survey (NHIS)
(continued)

Date or Frequency:	These data have been collected since 1969, though only data from 1982 through the present are currently available for download from the Internet. The most recent, complete file available is from 2005.
Data Collection Methodology:	NHIS is currently conducted via a personal household interview with a knowledgeable adult household representative using computer-assisted personal interviewing (CAPI) technology.
Participation:	Optional, without incentives
Response Rate:	For 2003, the overall response rate was 89.2 percent for a sample of 35,921 households (92,148 individuals). The Sample Adult subsample had a response rate of 74.2 percent, while the Sample Child subsample had a response rate of 81.1 percent.
Sampling Methodology:	NHIS utilizes a stratified, multi-stage probability design to reflect the overall noninstitutionalized U.S. population. The sample is drawn from a geographic frame designated using the most recent decennial Census. Names and addresses are derived in a separate listing activities conducted specifically for NHIS. From 1995 through 2005, African American and Hispanic households were oversampled in order to facilitate better estimates for these populations. Beginning in 2006, households with at least one Asian member are also oversampled.
Analysis:	Equations calculating standard errors can be found in the Survey Description document on page: www.cdc.gov/nchs/about/major/nhis/quest_data_related_1997_forward.htm.
Strengths:	Data set contains a large number of AI/AN/NA respondents. Data are collected on key policy issues, including health and child welfare. There are multiple years of data available. Certain years can be linked to the Medical Expenditure Panel Survey (MEPS) to produce a very rich database that includes medical care utilization data.
Limitations:	The NHIS data conform to the revised OMB guidelines, which mean that data for the Asian population is collected separately from that of the Native Hawaiian and other Pacific Islander (NH/PI) population. NCHS data confidentiality standards do not permit the release the population counts for the NH/PI population separately because of the small size of this population. This group is combined with other small population groups into an aggregated group labeled "Race group not releasable" on the public use data file. Researchers wishing to do analysis on the NH/PI population can submit a proposal to the NCHS Research Data Center; however there is a charge which depends on the amount of time and help needed. Information on this process can be found at: http://www.cdc.gov/nchs/r&d/rdc.htm

National Health Interview Survey (NHIS)
(continued)

Access Requirements and Use Restrictions:	Data are available to the public at no cost.
Contact Information:	National Center for Health Statistics, Hyattsville, MD 20782 (301) 458-4000 or (866) 441-NCHS http://www.cdc.gov/nchs/nhis.htm

National Hospital Ambulatory Medical Care Survey (NHAMCS)

Sponsor: U.S. Department of Health and Human Services (DHHS)/Centers for Disease Control and Prevention (CDC)/National Center for Health Statistics (NCHS)

Description: The National Hospital Ambulatory Medical Care Survey (NHAMCS) collects data on the utilization and provision of ambulatory care services in the emergency and outpatient departments of noninstitutional general and short-stay hospitals in the United States. Hospital staff complete a one-page questionnaire for each patient visit sampled during a 4-week reporting period. Collected data include hospital characteristics, patient demographic characteristics (age, sex, race, ethnicity), and visit characteristics (patients' symptoms, complaints or other reasons for the visit, physicians' diagnoses, diagnostic and therapeutic services ordered or provided at the visit including medications, expected sources of payment, visit disposition). Excluded from the sample are federal, military, and Veterans Administration hospitals. Sample data must be weighted to produce national estimates.

Relevant Policy Issues: Measurement of Health Status, Key Health Disparities, and Factors Contributing to Measured Health Disparities.

Data Type(s): Survey

Unit of Analysis: Patient visit

Identification of AI/AN/NA: Hospital staff are asked to report one or more races (up to 5) for each sampled visit. The public use data files include five single race categories and an aggregated category for visits with more than one race checked.

- White
- Black/African American
- Asian
- Native Hawaiian/Other Pacific Islander (NH/PI)
- American Indian/Alaska Native (AI/AN)
- More than one race reported

AI/AN/NA Population in Data Set: Total number of records in 2004 Emergency Department data set: 36,589
Total number of records in 2004 Outpatient Department data set: 31,783
AI/AN: 209 in 2004 Emergency Department data set; 149 in 2004 Outpatient Department data set

- NH/PI: 213 in 2004 Emergency Department data set; 305 in 2004 Outpatient Department data set

Geographic Scope: The geographic scope of the study is national. Analysis is possible for the following regions: Northeast, Midwest, South, West.

Date or Frequency: Data are available annually from 1992.

National Hospital Ambulatory Medical Care Survey (NHAMCS)
(continued)

Aggregation: Each year of data has, on average, 120-170 outpatient department visits and 160-240 emergency department visits by persons reported as AI/AN only. In addition, each year has, on average, 140-300 outpatient department visits and 200-240 emergency department visits by persons reported as NH/PI only. According to the NCHS, researchers frequently combine years of data for analysis in order to achieve reliable estimates. Researchers considering aggregation should take special note of changes in sample design variables across the years, as these will affect variance estimation. They should also be particularly aware of any possible clustering by race that may affect sample estimates. The format and content of the survey questionnaires has also changed across the years. Data must be weighted to produce national estimates, and researchers may wish to seek guidance about the use of weights with aggregated files.

Data Collection Methodology: The U.S. Census Bureau acts as the data collection agent for the NHAMCS. Hospital staff are trained by Census field representatives to sample patients and to complete the 1-page reporting form for each sampled visit.

Participation: Optional, without incentives.

Response Rate: In the 2004 survey, the response rate for participating hospitals was 91 percent.

Sampling Methodology: NHAMCS utilizes a multistage probability sample design where geographic primary sampling units (PSUs) are selected in the first stage; a fixed panel of 600 hospitals, developed from the SMG Hospital Market Database in 1991 and updated using data products from Verispan, LLC, comprises the second stage; the selection of outpatient department clinics and emergency service areas (ESAs) from the outpatient and emergency departments of the sampled hospitals constitutes the third stage; and the selection of patient visits within sampled clinics and ESAs during a randomly selected 4-week reporting period is the fourth stage.

Analysis: The weighting procedure produces essentially unbiased national estimates and has three components: 1) inflation by reciprocals of the probabilities of selection, 2) adjustment for nonresponse, and 3) a population weighting ratio adjustment. Two data sets are produced – one for outpatient department visits and one for emergency department visits. Patient visit weights are provided on each file to produce accurate national estimates.

Strengths: Data are collected on key policy issues pertaining to health. There are multiple years of data available.

Limitations: In the 2004 survey, the overall item nonresponse rate is low; however it is 20.9 percent for ethnicity and 19.2 percent for race. On the outpatient department file, race and ethnicity were missing on 11.9 percent and 11.8 percent of records respectively. On the emergency department file, race and ethnicity were missing for 10.6 percent and 15.1 percent of records respectively. Race and ethnicity are imputed in both files by randomly assigning a value from another sampled visit

with similar characteristics. There are also relatively few visits by patients categorized as AI/AN or NH/PI.

Access Requirements and Use Restrictions:

Data are available to the public at no cost. Restricted files which contain additional variables and non-masked data can be accessed by applying to the NCHS Research Data Center and paying a fee.

Contact Information:

Main website for NHAMCS
http://www.cdc.gov/nchs/nhamcs.htm

Data Download:
http://www.cdc.gov/nchs/about/major/ahcd/ahcd1.htm#Micro-data

Contact Information:
National Center for Health Statistics
Ambulatory Care Statistics Branch
3311 Toledo Road, Rm. 3409
Hyattsville, MD 20782
(301) 458-4600

National Household Education Surveys Program (NHES)

Sponsor: U.S. Department of Education (DoE)/National Center for Education Statistics (NCES)

Description: The National Household Education Surveys Program (NHES) provides descriptive data on the educational activities of the U.S. population and offers researchers, educators, and policymakers a variety of statistics on the condition of education in the United States. The NHES surveys cover learning at all ages, from early childhood to school age to adulthood. The NHES uses a repeating cross-sectional design that allows for the study of trends related to educationally important topics.

While there are many surveys included in this system, this profile includes only the surveys that have more than 100 AI/AN/NA individuals across multiple years, and that were administered in 1995 or later (note that the first NHES survey was fielded in 1991). Surveys that are included in this profile are the Parent and Family Involvement in Education Survey (PFI-NHES), the Early Childhood Program Participation Survey (ECPP-NHES), the Adult Education Survey (AE-NHES) (also called the Adult Education and Lifelong Learning Survey (AELL-NHES)), the Before- and After-School Programs and Activities Survey (ASPA-NHES), and the 1999 Parent Survey (Parent-NHES, which includes some items from the PFI-NHES, ECPP-NHES, and ASPA-NHES).

Relevant Policy Issues: Educational Attainment, Educational Opportunities, and Factors Contributing to Educational Disparities.

Data Type(s): Survey

Unit of Analysis: Individual

Identification of AI/AN/NA: Race is self-reported on all the surveys. The questions asking about race were changed between the 2003 and 2005 collections in order to meet new OMB requirements for collecting this information. Specifically, the questions were changed to allow respondents to be classified as more than one specific race (e.g., as both White and Black). In previous years (i.e., 1995-2003), respondents were classified as either belonging to only one racial group or as being multiracial without those specific races being identified. For example, the 2005 Adult Education Survey included the following race question:

Which of the following races do you consider yourself to be? You may name more than one.

- White
- Black
- American Indian or Alaska Native (AI/AN)
- Asian
- Native Hawaiian or other Pacific Islander (NH/PI)
- Other specify

In 2001, the AE-NHES race questions was:

Are you...

- White
- Black
- American Indian or Alaska Native
- Asian or Pacific Islander
- Some other race? (Interviewers are instructed to use this category if no race is selected or if more than one race is selected.)

The race question in all surveys prior to 2005 was structured the same way as this question from the 2001 AE-NHES Survey. Note that also prior to 2005, the categories "Asian" and "Native Hawaiian or other Pacific Islander" were combined into a single category: "Asian or Other Pacific Islander."

For all NHES surveys, information about Hispanic ethnicity is collected in a separate question.

AI/AN/NA Population in Data Set:

PFI-NHES
1996: Out of 20,792 records, 231 are AI/AN
2003: Out of 12,426 records, 108 are AI/AN

ECPP-NHES
1995: Out of 14,064 records, 113 are AI/AN
2001: Out of 6,749 records, 51 are AI/AN
2005: Out of 7,209 records, 233 are AI/AN and 66 were NH/PI

AE-NHES and AELL-NHES
1995: Out of 19,722 records, 160 are AI/AN
1999: Out of 6,697 records, 51 are AI/AN
2001: Out of 10,873 records, 84 are AI/AN
2005: Out of 8,904 records, 355 are AI/AN and 51 were NH/PI

ASPA-NHES
2001: Out of 9,583 records, 79 were AI/AN
2005: Out of 11,684 records, 374 were AI/AN and 79 were NH/PI

Parent-NHES:
1999: Out of 24,600 records, 193 were AI/AN

Geographic Scope:

The geographic scope of NHES is national. While zip code was collected on the surveys, it is not included in the public use data files. The files do contain the variable ZIPURBAN, which identifies records as being inside or outside an urban region. Additionally, records contain a Census region with the four values: Northeast, South, Midwest, West.

Date or Frequency: The PFI-NHES survey was conducted in 1996 and 2003. The ECPP-NHES survey was conducted in 1995, 2001, and 2005. The AE-NHES survey was conducted in 1995, 1999, 2001 and 2005. In 2001, the AE-NHES survey was called the Adult Education and Lifelong Learning Survey (AELL-NHES). The ASPA-NHES survey was conducted in 2001 and 2005. The Parent-NHES was conducted in 1999. For all years of administration, the NHES data collection period runs from January to April.

The NHES will continue to be conducted regularly in the future, covering the same topics investigated in earlier collections. In 2007, surveys will cover Adult Education for Work-Related Reasons, School Readiness, and Parent and Family Involvement in Education.

Aggregation: It is possible to combine multiple years of data for the NHES surveys, but researchers should be aware that there is a sizable time difference between the administrations of some of the surveys (for example, the PFI-NHES survey was conducted in 1996 and 2003). In addition, in 1995, the NHES surveys switched the way in which they selected their sample (from using a modified Mitofsky-Waksberg method to a list-assisted method). For this reason, it is recommended that the samples from NHES surveys conducted prior to 1995 not be combined with samples later than 1995. As the NHES is weighted data, new weights would need to be developed in order to combine survey data across multiple years.

Researchers also should review any survey questions of interest before combining them to ensure the question text and response categories are comparable. Change in the race question text between 2003 and 2005 may affect the ability to aggregate 2005 data with data from previous years.

If researchers choose to combine data sets across multiple years, the end results would be:

PFI-NHES surveys from 1996 and 2003: Total AI/AN : 339
ECPP-NHES surveys from 1995, 2001 and 2005: Total AI/AN : 397
AE-NHES surveys from 1995, 1999, 2001 and 2005: Total AI/AN: 650
ASPA-NHES surveys from 2001 and 2005: Total AI/AN: 453

Data Collection Methodology: NHES is conducted as a random digit dial (RDD) telephone survey using computer-assisted telephone interviewing (CATI). Each household contact begins with a screener to obtain information used to sample adults and children for extended interviews (the topical surveys). Since 1996, in order to introduce the survey, advance letters have been sent out to all sample households where an address was obtained through a commercial address matching service.

Participation: Optional, with incentives

Response Rate:

Response rates:
1996 PFI-NHES: Weighted response rate of 62.5%
2003 PFI-NHES: Weighted response rate of 53.8%

1995 ECPP-NHES: Weighted response rate of 66.3%
2001 ECPP-NHES: Weighted response rate of 59.9%
2005 ECPP-NHES: Weighted response rate of 56.4%

1995 AE-NHES: Weighted response rate of 58.6%
1999 AE-NHES: Weighted response rate of 62.3%
2001 AELL-NHES: Weighted response rate of 53.4%
2005 AE-NHES: Weighted response rate of 47.6%

2001 ASPA-NHES: Weighted response rate of 59.7%
2005 ASPA-NHES: Weighted response rate of 56.3%

1999 Parent-NHES: Weighted response rate of 66.7%

The NHES data user manuals provide weighted response rates as these rates give a better description of the success of the survey with respect to the population of interest. The response rate indicates the percentage of possible interviews that have been completed, taking all sampling stages into account. The weighted response rate is similar to the unweighted response rate unless the probabilities of selection vary considerably.

Sampling Methodology:

Beginning in 1995, the NHES surveys began using a list-assisted method to select the random digit dial (RDD) sample. With the list-assisted method, an equal probability random sample of telephone numbers is selected from all telephone numbers that are in 100-banks (numbers in a 100-bank have the same first 8 digits of the 10-digit telephone number) in which there is at least one residential telephone number listed in the white pages directory (the listed stratum). Both listed and unlisted telephone numbers are included in the listed stratum. Telephone numbers in 100-banks with no listed telephone numbers (the zero-listed stratum) were not sampled.

Additionally, a within-household sampling scheme was developed to limit the number of persons sampled for extended interviews in each household in order to reduce respondent burden.

Analysis:

Design effects (deff):
1996 PFI-NHES survey deff = 1.3 for the Full Sample
2003 PFI-NHES survey deffs = 1.3 for the Full Sample, and 1.4 for the race/ethnicity subgroups

1995 ECPP-NHES survey deff = 1.2 for the Full Sample
2001 ECPP-NHES survey deff = 1.2 for the Full Sample
2005 ECPP-NHES survey = 1.4 for the Full Sample, 1.3 for Preschoolers

National Household Education Surveys Program (NHES)
(continued)

1995 AE-NHES survey deff = 1.3 for the Full Sample
1999 AE-NHES survey deff = 1.3 for the Full Sample, 1.4 for Participants
2001 AELL-NHES survey deff = 1.3 for the Full Sample
2005 AE-NHES survey deff= 1.6 for the Full Sample

2001 ASPA-NHES survey deff = 1.3 for the Full Sample
2005 ASPA-NHES survey deff = 1.3 for the Full Sample

1999 Parent-NHES survey deff = 1.3 for the Full Sample

Strengths: Multiple years of data are available. Documentation is detailed and comprehensive.

Limitations: Most of these studies contain a relatively small number of AI/AN/NA respondents. While it may be possible to combine the samples from multiple years of the surveys, this is a complicated procedure. One of the biggest issues related to aggregation is the variation in the race questions over the years.

Access Requirements and Use Restrictions: Data are available online at no cost.

Contact Information: The NHES data can be downloaded via the Internet at: http://nces.ed.gov/nhes/dataproducts.asp

The NHES staff can be contacted by sending an email to: nhes@ed.gov.

National Household Travel Survey (NHTS)

Sponsor: U.S. Department of Transportation/Federal Highway Administration, Bureau of Transportation Statistics, and National Highway Traffic Safety Administration

Description: The 2001 National Household Travel Survey (NHTS) was developed to gather comprehensive data on travel and transportation patterns in the United States. Data were collected on daily trips taken in a 24-hour period, as well as long distance trips collected for the 4-week period prior to the travel day. Data collected for the daily trips include the purpose of the trip, the means of transportation used, how long the trip took, when the trip took place, and characteristics of private vehicle ownership. The survey measured personal travel, which excluded travel made as part of the respondent's job (i.e., business trips). The NHTS data are used primarily for gaining a better understanding of travel behavior. For example, NHTS data are used to quantify travel behavior, analyze changes in travel characteristics over time, relate travel behavior to the demographics of the traveler, and study the relationship of demographics and travel over time. Additionally, people in fields outside of transportation use the NHTS data. For example, social service agencies may use the data to learn more about how low-income households currently meet their travel needs.

In addition to the national sample, planning organizations could purchase add-on surveys in their state or specific county so that reliable estimates could be made for that geographic area. The nine add-on samples include the following areas:

- State of Wisconsin
- State of New York
- State of Texas
- Hawaii (state-wide, excluding Oahu)
- Oahu, Hawaii
- Baltimore, Maryland
- Des Moines, Iowa
- Lancaster, Pennsylvania
- Edmonson, Carter, Pulaski, and Scott Counties, Kentucky

Relevant Policy Issues: Transportation Availability.

Data Type(s): Survey

Unit of Analysis: The NHTS data are organized into five different data files and contain data on the 26,038 households from the national sample, and the 43,779 completed add-on households. There were over 160,000 person interviews from the national sample and add-on households, with about 100,000 of them from the add-on sample. Records from each data file can be linked to one another using the Household ID number. Descriptions for each data file are as follows:

National Household Travel Survey (NHTS)
(continued)

Household file: Contains one record per sampled household. Variables include the number of vehicles, type of residence, household income, and information on the primary household respondent.

Person file: Contains one record per person who completed a person interview. Variables include information on traveling to work; the number of miles driven; customer satisfaction with transportation arrangements; and person demographics such as age, race, driver status, and medical condition.

Vehicle File: Contains one record per each vehicle owned, leased, or available for regular use by the household members in each sample household. Variables include type of vehicle, vehicle ownership, mileage, and housing characteristics.

Travel day trip file: Contains one record per each trip taken by an interviewed person in a sampled household, for the household's randomly- assigned travel day.

Long trip file: Contains one record per each trip of 50 miles or more away from home. The long distance trip data were collected in the national sample and NY and WI add-on samples only.

For the NHTS study, household members include all people who think of the sampled household as their primary place of residence. It includes persons who usually stay in the household but are temporarily away on business, vacation, or in a hospital. It does not include people just visiting, such as college students who normally live away at school.

Identification of AI/AN/NA:

Race is self-reported, using the following survey item:

I'm going to read a list of races. Please tell me which best describes your race. Are you…
(Interviewers are instructed to code all that apply.)

- White
- African American, Black
- Asian
- American Indian, Alaskan Native (AI/AN)
- Native Hawaiian, or other Pacific Islander (NH/PI)

The following categories are not asked but are coded in the data files:

- Multiracial
- Hispanic/Mexican
- Other
- Refused
- Don't Know

National Household Travel Survey (NHTS)
(continued)

AI/AN/NA Population in Data Set:

2001 NHTS Household File
Total number of records: 69,817 households
AI/AN primary household respondent: 401
NH/PI primary household respondent: 370
White/AI primary household respondent: 547
AI/Hispanic primary household respondent: 44

2001 NHTS Person File
Total number of records: 160,758 completed interviews
AI/AN person in household: 882
NH/PI person in household: 1,027
White/AI person in household: 1,275
AI/Hispanic person in household: 115

2001 NHTS Vehicle File
Total number of vehicles: 139,382
AI/AN primary driver: 697
NH/PI primary driver: 720
White/AI primary driver: 1,189
AI/Hispanic primary driver: 62

2001 NHTS Day Trip File
Total number of day trips: 642,292
AI/AN traveler: 3,383
NH/PI traveler: 3,794
White/AI traveler: 4,834
AI/Hispanic traveler: 413

2001 NHTS Long Trip File
Total number of long trips: 45,165
AI/AN traveler: 277
NH/PI traveler: 127
White/AI traveler: 609
AI/Hispanid traveler: 24

Geographic Scope:

The geographic scope of the study is national. Geographic indicators included are state, Census division, Census region, and Metropolitan Statistical Area (MSA) or Consolidated Metropolitan Statistical Area (CMSA).

The NHTS data are supplemented with tract and block group data that are derived from the 2000 Census data. These data are used to describe the characteristics of the areas where the NHTS respondents live, including descriptors such as housing units per square mile, an urban/rural code, and population density per square mile.

Date or Frequency:

The NHTS interviews were conducted from April 2001 through May 2002.

National Household Travel Survey (NHTS)
(continued)

The NHTS resulted from integrating two national travel surveys: the Federal Highway Administration-sponsored Nationwide Personal Transportation Survey (NPTS) and the Bureau of Transportation Statistics-sponsored American Travel Survey (ATS). The NPTS collected detailed information on personal travel patterns using daily travel surveys, and was conducted in 1969, 1977, 1983, 1990 and 1995. The ATS obtained information about long-distance travel of persons living in the United States and was collected in 1977 and 1995.

Data Collection Methodology: Data collection consisted of three main phases. First, a household interview was conducted using computer assisted telephone interviewing (CATI) technology. The household interview was designed to collect information about the household, household members, and vehicles available to the household, and to elicit participation in the travel diary task. Next, travel diaries were mailed to the households. Each household in the sample was assigned a specific 24-hour "Travel Day" and kept diaries to record all travel by all household members for the assigned day. Respondents were also asked to document a 28-day "Travel Period" in order to collect longer-distance travel (over 50 miles from home) for each household member, including information on long commutes, airport access, and overnight stays. The assigned travel day was the last day of the assigned travel period. For a household to be included in any of the data sets, interviews had to be completed with at least half of the household adults. Finally, for the national sample and the New York and Wisconsin add-on samples, odometer readings from the household vehicles were collected from the respondent.

Participation: Optional, with incentives.

Response Rate: Overall weighted response rates were 41.2 percent for the national sample and 38.9 percent for the full sample (includes national sample and 9 add-on samples). These response rates were an improvement from the previous 1995 survey response rates, and are considered high for travel surveys of this type.

Sampling Methodology: NHTS collected travel data from the civilian, non-institutionalized population of the United States. Sampling was done by creating a random-digit dialing (RDD) list of telephone numbers. The sampling frame consisted of all telephone numbers in "100-banks" of numbers in which there was at least one listed residential number. (Each 100-bank contains the 100 telephone numbers with the same area code, exchange, and first two of the last four digits of the telephone number.)

Analysis: The NHTS is a weighted data set. The weights reflect the selection probabilities and adjustments to account for nonresponse, undercoverage, and multiple telephones in a household. To obtain estimates that are minimally biased, weights must be used. Tabulations without weights may be significantly different than weighted estimates and may be subject to large bias. There are separate sets of weights for the full sample and for the national sample. For each set, there are household weights, person weights, travel day and travel period weights. The NHTS methodology report describes the process for applying the weights appropriately.

National Household Travel Survey (NHTS)
(continued)

Strengths: There are multiple years of data available. Documentation for the NHTS is readily available, very detailed, and extremely clear.

Limitations: During the 2001 NHTS data collection period, the September 11th terrorist attacks occurred and severely disrupted travel in the United States for months. These attacks altered the amount and modes of travel that were being documented during this data collection.

Although there are multiple years of data available for comparison purposes, aggregation of the data is not recommended. The sample design varied across administrations of the survey. Also, there were changes in national travel behavior across the years of data collection due to the state of the economy, the price of oil, and the terrorist attacks of September 11, 2001. Some significant variations can be expected in travel trend analysis, but the nature of these variations is such that aggregation of the data across multiple years is discouraged.

As the data from the NHTS study is organized into four separate files, it may require strong programming skills to merge the files together if this is required for a specific research questions.

Access Requirements and Use Restrictions: The public use data are available to the public at no cost.

Contact Information:
Federal Highway Administration/DOT
The Office of Highway Policy Information (HPPI)
Rm 3306
Washington, DC 20590
Phone: (202) 366-5021
Fax: (202) 366-7742

The NHTS public use database can be downloaded directly from:
http://nhts.ornl.gov/2001/html_files/download_directory.shtml.

National Indian Education Study (NIES)

Sponsor: U.S. Department of Education (ED)/National Center for Education Statistics (NCES)/Office of Indian Education (OIE)

Description: The National Indian Education Study (NIES) focuses on both the academic achievement and educational experiences of fourth and eight grade students across the country. This activity is a collaborative effort among Indian tribes and organizations, the Bureau of Indian Affairs (BIA), and state and local education agencies. Part One of the NIES is an augmentation of the National Assessment of Educational Progress (NAEP) 2005 reading and mathematics assessment to increase the representation of Bureau of Indian Affairs schools in the NAEP data and allow for separate analyses. Part Two is a separate survey focusing on issues of Indian education, such as the role of Indian culture in education. At the time this catalog was being compiled, further information concerning Part Two of the NIES had not been released. The results of the NIES are intended to assist policymakers, educators, and community members in making informed decisions to improve education of all American Indian and Alaska Native students.

Relevant Policy Issues: Educational Attainment.

Data Type(s): Survey

Unit of Analysis: Individual

Identification of AI/AN/NA: Race/ethnicity is collected from two sources: school records and student self-reports. The primary source of race/ethnicity data is based on the race reported by the school. Schools that are sampled for the study are asked to provide lists of all students in grades 4 and 8, along with basic demographic information, including race/ethnicity.

When school-recorded information is missing, student-reported data are used to determine race/ethnicity. All students who complete an assessment are asked some general student background questions, including questions about their race/ethnicity. Separate questions are asked about students' Hispanic ethnic background and about students' race.

Based on the school records and self-reports, students are categorized into one of the following mutually exclusive categories:

- White (non-Hispanic)
- Black (non-Hispanic)
- Hispanic
- Asian/Pacific Islander
- American Indian (including Alaska Native) (AI/AN)
- Unclassified

Unclassified students are those whose school-reported race was recorded as "other" or "unavailable" or was missing, or who self-reported more than one race category (i.e., "multi-racial") or none. Hispanic students may be of any race.

AI/AN/NA Population in Data Set: In the full 2005 NAEP reading assessment, a sample of about 166,000 fourth-grade and 159,000 eighth-grade students participated in the study. Of these, approximately 3,800 AI/AN students in grade 4 and approximately 3,400 AI/AN students in grade 8 are included in the NIES study. These numbers have been rounded to the nearest 100.

In the full 2005 NAEP mathematics assessment, a sample of about 172,000 fourth-grade and 162,000 eighth-grade students participated in the study. Of these, approximately 3,900 AI/AN students in grade 4 and approximately 3,500 AI/AN students in grade 8 are included in the NIES study. These numbers have been rounded to the nearest 100.

Geographic Scope: The geographic scope of the study is national. In addition to the national sample, data for American Indian/Alaska Native students can be analyzed for five regions of the country. These regions are Atlantic, North Central, South Central, Mountain, and Pacific. Furthermore, NIES focused on states with relatively high proportions of American Indian and Alaska Native students. The seven states included are Alaska, Arizona, Montana, New Mexico, North Dakota, Oklahoma, and South Dakota. Data concerning AI/AN students can be analyzed separately for each of these states, but comparisons between AI/AN students and students who are not AI/AN cannot be made at the state level.

Date or Frequency: NIES data were collected for the first time in 2005. Future administrations of the study are planned.

Data Collection Methodology: The reading and mathematics assessments are administered to participating students at the school location in paper and pencil format.

Participation: Optional, without incentives

Response Rate: The NIES student-weighted response rates are below:
Grade 4 reading: 93%
Grade 4 mathematics: 93%
Grade 8 reading: 91%
Grade 8 mathematics: 88%

Sampling Methodology: The NIES sample was designed as an augmentation of the 2005 NAEP reading and mathematics assessment samples of AI/AN students in the fourth and eighth grades. In past NAEP samples, BIA schools were identified as part of the national sample, and the resulting number of participating schools was usually small, fewer than five per grade. In order to create the NIES study in 2005, BIA schools were sampled as a part of each state sample, at the same rate as public schools in a given state. Therefore a BIA student had the same probability of selection as a

public school student in the same state. As a result, about 30 BIA schools were included per grade, thereby increasing the number of AI/AN students in the sample.

The national and regional results of the NIES include AI/AN students in all schools (public, private, Department of Defense, and BIA), while the state results are based on samples of AI/AN students in public and BIA schools only; however, the percentage of AI/AN students who are enrolled in schools other than public and BIA schools nationally is very small (between 1 and 2 percent, unweighted).

Authorization: Authorization for this study falls under the Executive Order 13336: American Indian and Alaska Native Education. This Executive Order is a follow-up to the No Child Left Behind Act.

Strengths: Data sets contain a large number of AI/AN/NA respondents. Data are collected on a key policy issue, education.

Limitations: No major limitations were identified with Part One of the study. Details from Part Two were not available at the time this catalog was compiled and thus cannot be included in this profile.

Access Requirements and Use Restrictions: The data will be made available to researchers with NCES license. There is no cost. The steps for obtaining a license are detailed at the following website: http://nces.ed.gov/pubsearch/licenses.asp.

Contact Information:
Office of Indian Education
U.S. Department of Education
400 Maryland Avenue SW 5C132
Washington, DC 20202-6335
Phone: (202) 260-7485
Fax: (202) 260-7779
E-mail: indian.education@ed.gov

Specific questions about the NIES to can be directed to Jeff Johnson (jeff.johnson@ed.gov) or Taslima Rahman (taslima.rahman@ed.gov).

National Longitudinal Mortality Study (NLMS)

Sponsor:	U.S. Department of Health and Human Services (DHHS)/National Institutes of Health/National Cancer Institute (NCI); National Heart, Lung, and Blood Institute (NHLBI); National Institute on Aging (NIA); National Center for Health Statistics (NCHS); and the U.S. Census Bureau
Description:	The National Longitudinal Mortality Study (NLMS) examines the effects of demographic and socio-economic characteristics on differentials in mortality. The NLMS is a unique research database in that it is based on a random sample of the non-institutionalized population of the United States. Records from the Annual Social and Economic Supplement, and the Census Current Population Surveys (CPS) are matched to mortality information from death certificates available for deceased persons through NCHS. Extensive demographic, social, economic, and occupation information is collected in the CPS. The study currently consists of approximately 2.3 million records with over 250,000 identified mortality cases.
Relevant Policy Issues:	Measurement of Health Status, Key Health Disparities, and Factors Contributing to Measured Health Disparities.
Data Type(s):	Registry
Unit of Analysis:	Individual
Identification of AI/AN/NA:	Most of the race information in the NLMS is based on the CPS.
AI/AN/NA Population in Data Set:	Based on frequencies provided for each file (for years 1979 through 1998) in the public use data file reference manual, there are 19,779 American Indians/Alaska Natives (AI/AN) and 1,504 Native Hawaiians/Other Pacific Islanders (NH/PI) in this file. For years 1992 through 1998, the average number of AI/AN in the data files are 848 and 220 for NH/PI persons.
Geographic Scope:	The geographic scope of the study is national. Geographic indicators available for analysis include region, state, county, urban status, and SMA status.
Date or Frequency:	Data files linking Census data to death certificate information are available for 1973-2002.
Data Collection Methodology:	Census data are linked to mortality information obtained from death certificates available for deceased persons through the NCHS.
Participation:	This is a secondary data linkage and does not require participation by individual respondents.
Strengths:	Data are collected on key policy issues, including health. There are multiple years of data available. The linkage of the individual social and economic data with the mortality outcomes provides the resource for extensive analysis.

National Longitudinal Mortality Study (NLMS)
(continued)

Limitations: The study is based on specific survey months of the CPS, the Annual Social and Economic Supplement, and a subset of the 1980 Census. These are one-time data collection processes with no subsequent data collection. Therefore, one limitation of NLMS data is that they provide a one-time only baseline measurement of subjects in a long-term follow-up situation. Another limitation of these data is that, although the CPS and Census instruments provide extensive data collection capabilities in specific subject matter areas, desirable general or specific health information is not collected and smoking status is available only on a limited number of records.

Access Requirements and Use Restrictions: A public use data file is available. Potential users must submit a data use agreement to the NHLBI. Data use agreement form can be found at:

http://www.census.gov/nlms/docs/form.doc.

Research access to the entire NLMS database may be arranged through the principal investigators of the NLMS sponsoring agencies.

Contact Information: Project Main Website: http://www.census.gov/nlms/index.html

Norman J. Johnson
U.S. Census Bureau
4700 Silver Hill Road
DSMD, Room 3725-3
Suitland, MD 20746
Ph: (301) 763-4270
FAX: (301) 457-3766
email: norman.j.johnson@census.gov

Reports of Interest: The NLMS website provides a list of published articles based all or in part on either the full NLMS database or the NLMS public-use file. The URL for this bibliography is: http://www.census.gov/nlms/bibliography.html.

National Mortality Followback Survey (NMFS)

Sponsor:	U.S. Department of Health and Human Services (DHHS)/Centers for Disease Control and Prevention (CDC)/National Center for Health Statistics (NCHS)
Description:	The National Mortality Followback Survey Program (NMFS), begun in the 1960s by NCHS, draws a sample of U.S. residents who die in a given year and supplements their death certificate information with information from the next of kin or another person familiar with the decedent's life history. This information, sometimes enhanced by administrative records, provides a unique opportunity to study the etiology of disease, demographic trends in mortality, and other health issues.
Relevant Policy Issues:	Measurement of Health Status and Key Health Disparities.
Data Type(s):	Survey
Unit of Analysis:	Individual
Identification of AI/AN/NA:	In the 1993 data, death certificate information is coded into the following categories: • White • Black • American Indian, Eskimo, Aleut (AI/AN) • Asian and Pacific Islander • Could not match
AI/AN/NA Population in Data Set:	In the 1993 NMFS data, out of 22,957 total unweighted cases, 205 unweighted cases are identified as AI/AN.
Geographic Scope:	The geographic scope of the study is national. (South Dakota did not participate in 1993.) Analysis (in the 1993 data) is also possible by region and by size of county.
Date or Frequency:	The 1993 NMFS is the sixth in a series of surveys, first initiated by NCHS in the early 1960's. Data are available for both the 1993 NMFS and the 1986 NMFS.
Data Collection Methodology:	The sampling frame for the 1993 NMFS is the 1993 Current Mortality Sample (CMS). The CMS is a 10 percent systematic random sample of states' death certificates. The proxy respondent questionnaire is completed through a telephone or in-person interview.
Participation:	Optional, without incentives
Response Rate:	The overall response rate was 83 percent.

National Mortality Followback Survey (NMFS)
(continued)

Sampling Methodology: The sampling frame for the 1993 NMFS is the 1993 Current Mortality Sample (CMS). The CMS is a 10 percent systematic random sample of states' death certificates. A sample of 22,957 death certificates from the Current Mortality Sample was drawn. The sample was selected by broad age groups (15 years or older), two racial groups (black vs. nonblack), and gender within 12 causes of death (suicide; homicide; injuries to motor vehicle drivers, pedestrians, and motorcycle owner operators; other motor vehicle injuries; non-motor vehicle injuries; HIV; cancer; chronic obstructive pulmonary disease; heart disease; alcohol abuse; drug abuse; and all other causes). In order to produce more robust analysis, black decedents, certain causes of death, and certain age groups were oversampled (45.5 percent of all cases).

Analysis: To produce national estimates, researchers must use weights and adjust for the complex sampling design for all analyses.

Authorization: The 1993 NMFS was conducted under the authorization of the Public Health Service Act (Title 42, United States code, Section 242k).

Strengths: Data are collected on key policy issues including health status and health disparities. Documentation is thorough and available online.

Limitations: The sample size for AI/AN is small.

Access Requirements and Use Restrictions: The 1986 and 1993 data from the NMFS are available on CD-ROM at no charge. The 1993 NMFS data are also available by directly downloading from the CDC website.

Contact Information: For questions concerning the 1993 National Mortality Followback Survey or to obtain a CD-ROM, contact:

Mortality Statistics Branch
Division of Vital Statistics
National Center for Health Statistics
Centers for Disease Control and Prevention
3311 Toledo Road, Room 7318
Hyattsville, Maryland 20782
(301) 458-4666

The data can be directly downloaded from:
http://www.cdc.gov/nchs/about/major/nmfs/nmfs.htm.
Documentation for the 1993 survey is also available at this website.

National Survey of America's Families (NSAF)

Sponsor: The Urban Institute

Description: The National Survey of America's Families (NSAF) is part of The Urban Institute's Assessing the New Federalism project. Its purpose is to track the effects of recent federal policy changes decentralizing many social programs, and to provide a comprehensive look at the well-being of children and non-elderly adults. The survey provides quantitative measures of child, adult and family well-being in America, with an emphasis on persons in low-income families. The survey gathers data on economic, health and social characteristics of children and families in order to estimate well-being. Specific topics include: participation in government programs; employment; earnings and income; economic hardship; educational attainment; training; family structure; housing arrangements; health insurance coverage; access to and use of health services; health status; psychological well-being; participation in religious and volunteer activities; knowledge of social services; and attitudes about work, welfare, health care and childbearing. In 2002, interviews were conducted with more than 40,000 families, yielding information on more than 100,000 people. Earlier rounds of the survey were conducted in 1997 and 1999.

Relevant Policy Issues: Measurement of Health Status, Factors Contributing to Measured Health Disparities, Income Status, Economic Assistance Program Participation Rates, Economic Opportunity, Housing Quality, and Measure of Child Well Being and Child Care arrangements.

Data Type(s): Survey

Unit of Analysis: Individual or family level.

Identification of AI/AN/NA: What is (your/NAME's) race?
(Interviewers are instructed to probe by reading categories if necessary and if the respondent says "Native American," to verify by asking: "I am recording this as 'American Indian'—is that right?")

- White
- Black
- American Indian, Aleutian, or Eskimo (AI/AN)
- Asian/Pacific Islander
- Other (Specify)

The race question was identical in the 1999 study. In 1997, the race question did not include the follow-up probe for "Native American."

AI/AN/NA Population in Data Set: There are four person-level data files available for public use. These files include the Adult Pair, Random Adult, Childless Adult, and Focal Child. Additionally, there are two family-level files that do not contain race information, but can be merged with the individual-level data files.

The 2002 Focal Child File contains one record for each child (ages 0 – 17) included in the NSAF study. Out of 34,332 completed interviews in this file, 491 respondents are identified as American Indian/Native American/Aleutian or Eskimo.

There are also three data files for the adult respondents included in the NSAF. In the survey, some questions are asked about both the respondent and his or her spouse or partner (if one exists), others are asked about either the respondent or his or her spouse or partner (randomly chosen), and still others are asked about only the respondent.

The 2002 Adult Pair File contains records for both the respondent and his/her spouse or partner. Out of 70,577 completed interviews in this file, 1,018 individuals are identified as American Indian/Native American/Aleutian or Eskimo.

The 2002 Random Adult File contains data elements from the extended interview that are specific to a randomly selected adult (this is a subset of the Adult Pair File). The random adult data set contains variables based on questions asked during the extended interview that were specific to a randomly selected adult (either the respondent or the spouse/partner). This situation occurs only in sections E (Past Year Health Insurance Coverage) and F (Health Care use and Access) of the NSAF questionnaire. Out of 49,507 completed interviews in this file, 752 respondents are identified as American Indian/Native American/Aleutian or Eskimo.

The 2002 Childless Adult File contains data elements representing households without children, where up to two childless adults between the ages of 18 and 64 were selected for interviewing. Out of 15,279 completed interviews, 248 respondents are identified as being American Indian/Native American/Aleutian or Eskimo. This file is not available on the NSAF Online Statistical Analysis webpage, but can be downloaded seperately at: http://anfdata.urban.org/drsurvey/login.cfm

Geographic Scope:

Geographic indicator variables on the public use data files include a state indicator (including D.C.), the 5-digit Federal Information Processing Standards (FIPS) county code for counties with more than 250,000 persons in the 13 NSAF focal states, and the 4 Census regions of Northeast, Midwest (formerly North Central), West, and South.

The NSAF was designed to produce national estimates of the population under 65. Additionally, state estimates are possible for the 13 states that contained oversamples: Alabama, California, Colorado, Florida, Massachusetts, Michigan, Minnesota, Mississippi, New Jersey, New York, Texas, Washington, and Wisconsin. There is also a public use NSAF data set available for the California NSAF, conducted in 1997, 1999, and 2002 (containing 113 AI/AN individuals).

National Survey of America's Families (NSAF)
(continued)

Date or Frequency: The 2002 study was conducted from February to November. Previous rounds of data collection include the 1997 and 1999 study. At this time there are no plans for future rounds of data collection.

Aggregation: Researchers who wish to aggregate data across the three years of data collection for the NSAF study should first examine all survey items of interest to make certain the question text and the response options are identical. Although every effort was made to keep question wording unchanged between the rounds, there were some improvements made to some questions. Researchers should also examine the design of each wave very carefully. Changes were made in each wave that could affect survey weights and estimation; for example, in the third round of data collection, the sample size for nontelephone households in the study areas was reduced.

The number of AI/AN respondents for the 1997 and 1999 NSAF that could be included in an aggregation are below:

<u>AI/AN Respondents in 1999:</u>
Child file: 488
Adult pair file: 1,001
Random adult file: 721

<u>AI/AN Respondents in 1997:</u>
Child file: 529
Adult pair file: 971
Random adult file: 738

Data Collection Methodology: All interviews were conducted on the telephone by interviewers working in central interviewing facilities, using computer-assisted telephone interviewing (CATI) technology. In-person interviewers used cellular telephones to connect respondents in nontelephone households to the interviewing centers for the CATI interview.

Participation: Optional, with incentives

Response Rate: 2002: overall child response rate = 55.1%
2002: overall adult response rate = 51.9%

1999: overall child response rate = 62.4%
1999: overall adult response rate = 59.4%

1997: overall child response rate = 65.1%
1997: overall adult response rate = 61.8%

Sampling Methodology: The sample is representative of the civilian, non-institutionalized population under age 65. As with the prior two rounds of data collection (conducted in 1997 and 1999), the 2002 survey included oversize samples drawn in 13 states (listed

under Geographic Scope) to allow for the production of reliable estimates at the state level. The oversize state samples are supplemented with a balance of the United States sample to allow the creation of estimates at the national level as well.

The sampling frame consisted of a list-assisted, random-digit dialing (RDD) sample of telephone numbers supplemented by an area probability sample of nontelephone households. A short screening interview was used to identify and sample households based on age composition and household income. Once household eligibility was sampled, subsequent questions were asked to identify the children (age 0 to 17) or adults (age 18 to 64) in the household. Once this list was compiled, the CATI program sampled up to two children or up to two adults for subjects on the extended interview. If children were sampled, a series of questions was asked to determine the name and relationship of the person most knowledgeable about the selected child or children (the most knowledgeable adult).

Analysis: There are a series of methodology reports that accompany the NSAF public use data. Detailed information on calculating design effects (DEFF) and standard errors can be found in the following reports in the *2002 NSAF Variance Estimation, Report No. 4*, located at http://www.urban.org/UploadedPDF/900716_2002_Methodology_4.pdf; and the *NSAF Public Use File User's Guide, Report No. 11*, located at http://www.urban.org/UploadedPDF/900760_2002_Methodology_11.pdf.

Child, all races: DEFF = 1.77, effective sample size = 10,828
Adult, all races: DEFF = 2.13, effective sample size = 15,015
(Separate design effects and effective sample sizes were calculated for the Black and Hispanic populations, but not the AI/AN population.)

Strengths: Data are collected on many key policy issues, including family well-being measures such as assistance receipt, socioeconomic status and education. There are multiple years of data available.

For most survey questions the item nonresponse rates were very low, often less than 1 percent. For survey items with significantly higher levels of item nonresponse (such as income), missing responses were imputed using a standard "hot deck" method, for all three rounds of data collection.

Low-income families were oversampled. Even with this oversampling method, the NSAF contains a relatively small average margin of error for state-level estimates of low-income children and adults.

Limitations: No major limitations were identified.

National Survey of America's Families (NSAF)
(continued)

Access Requirements and Use Restrictions:	To access the public use data, researchers must register with the Urban Institute's website and agree to their terms of confidentiality. The data are available at no cost.
Contact Information:	The Urban Institute Assessing the New Federalism Policy Center 2100 M Street, NW Washington DC 20037 (202) 261-5377 nsaf@ui.urban.org The public use data can be accessed: http://anfdata.urban.org/drsurvey/login.cfm. Additionally, researchers can perform web-based analysis of the NSAF survey data using the online analysis tool at: http://www.urban.org/center/anf/analysisprelogin.cfm.

National Survey of Child and Adolescent Well-being (NSCAW)

Sponsor: U.S. Department of Health and Human Services (DHHS)/Administration for Children and Families (ACF)

Description: The National Survey of Child and Adolescent Well-being (NSCAW) provides nationally representative longitudinal data concerning children who are at risk of abuse or neglect or are in the child welfare system. Two samples of children were selected for NSCAW: children who were the subject of child abuse or neglect investigations conducted by Child Protective Service agencies (CPS sample) and children who had been in out-of-home or foster care for approximately one year and whose placement had been preceded by an investigation of child abuse or neglect (LTFC sample). The information comes from first-hand reports from children, parents, and other caregivers, as well as reports from caseworkers, teachers, and data from administrative records. The data include information on child and family functioning and well-being, service needs and utilization, and agency- and system-level factors that are likely to be related to child and family outcomes. Child outcomes of interest include health and physical well-being, cognitive and school performance, mental health, behavior problems, and social functioning and relationships.

Relevant Policy Issues: Measurement of Health Status, Measures of Well-being for Families/households, and Measures of Well-being for Children.

Data Type(s): Survey

Unit of Analysis: Individual

Identification of AI/AN/NA: What race are you? (Interviewers are instructed to code all that apply.)

- American Indian or Alaska Native (AI/AN)
- Asian
- Black or African American
- Native Hawaiian or other Pacific Islander (NH/PI)
- White

This format is used to ascertain race of child, caretaker, and caseworker.

AI/AN/NA Population in Data Set: In Wave 1, there are a total of 5,504 children in the CPS study group. Of these, 341 are identified as AI/AN. In addition, there are a total of 727 in the LTFC study group; 47 are identified as AI/AN.

Asians, Native Hawaiians and Pacific Islanders are combined into a single category in the data set and can not be analyzed separately.

Geographic Scope: The geographic scope of the study is national. The type of county residence (rural vs. urban) is identified in the sampling frame. (Counties with greater than 50 percent urban area are classified as urban. Remaining counties are classified as

rural.) Address is collected for follow-up purposes but does not appear to be recoded into geographic variables. Based on the documentation available, researchers should be able to conduct analyses by rural vs. urban county.

Date or Frequency: This is a longitudinal study with four waves of data collection. Baseline data collection began in Fall 1999 and was completed in April 2001. After the baseline, three additional waves of data collection occurred at 12 months, 18 months, and 36 months post-baseline.

Data Collection Methodology: Interviews with parents or caregivers and children were conducted in-person using computerized personal interviewing techniques (CAPI) in private settings (e.g., the home). Field personnel collected physical measurements and observation data for infants and toddlers. Caseworker and agency interviews were also conducted in-person.

The CAPI instrument guided the child interview and prompted the field representative to administer the required developmental assessments in the designated order. When prompted, the field representative retrieved the assessment materials and administered the various activities appropriate for the child's age.

Participation: Optional, with incentives

Response Rate: Child Interview Response Rates:

	CPS Sample	LTFC Sample
Wave 1	64.2%	73.4%
Wave 2	86.7%	92.5%
Wave 3	86.6%	94.0%
Wave 4	85.3%	88.5%

Sampling Methodology: The children in the NSCAW CPS and LTFC samples were selected using a two-stage stratified sample design. At the first stage, the U. S. was divided into nine sampling strata. Within each of these nine strata, primary sampling units (PSUs), geographic areas that encompass the population served by a single child protective services agency, were randomly selected using a probability-proportionate-to-size procedure that gave a higher chance of selection to PSUs having larger caseloads. The same numbers of children were then sampled within each PSU.

Analysis: Because the NSCAW sample design is complex (e.g., unequally weighted, stratified, and clustered), standard errors computed using standard statistical procedures that assume a simple random sample will generally be too small. Special software that accounts for the complex sample design is needed in order to correctly estimate the standard errors. The User's Manual provides detailed guidance on the use of commercially available software packages such as SUDAAN, Stata, WesVar, and the SAS procedures SURVEYMEANS and

SURVEYREG to correctly estimate the standard errors taking into account the complex sample design.

Authorization: In the Personal Responsibility and Work Opportunities Reconciliation Act of 1996, Congress directed the Secretary of the Department of Health and Human Services to conduct a national study of children who are at risk of abuse or neglect or are in the child welfare system. Congress directed that the study include a longitudinal component that follows cases for a period of several years; collects data on the types of abuse or neglect involved, agency contacts and services, and out-of-home placements.

Strengths: Data are collected on key policy issues, including health and child welfare. There are multiple years of data available. The unique content of the data will be extremely useful to researchers and policy makers. The documentation is comprehensive. Extensive nonresponse analysis has been conducted to identify potential sources of bias in the collected data. Because of the complexity of this data source, ACF and its contractor have developed the NSCAW Data Delivery System to aid analysts and other data users, support the reduction of the data set to a manageable level, support various programming environments, and provide an electronic codebook including frequency distributions for the variables in the CPS and LTFC cohorts.

Limitations: Data access is strictly controlled and is not available to employees at child welfare agencies. In the LTFC, the number of AI/AN is very small. Because of their longitudinal nature, weighted analyses of these data will be complex.

Access Requirements and Use Restrictions: Two different versions of the NSCAW data are available. The General Use Data has identifying information and geographic detail removed and variables posing a risk of respondent disclosure have been recoded. The Restricted Release version has geographic detail and fewer variables have been recoded, but this version presents a higher risk to respondent confidentiality. It is, therefore, only made available to researchers who can justify a need for high level access and who are willing to follow additional application requirements.

For both versions of NSCAW, access is limited to researchers who agree to the terms and conditions contained in the Data Use License. Only faculty and non-student research personnel at institutions that have an Institutional Review Board/Human Subjects Review Committee are eligible to order the data. While access to both versions of the NSCAW data require approval by an Institutional Review Board at the researcher's institution and close oversight by NDACAN in the form of a legally-binding licensing agreement, access to the Restricted Release Data also requires preparation of an application and data protection plan as well as willingness to cooperate with unannounced on-site inspections of the research facility.

University students may gain access to the NSCAW only as research staff who have been added to the project, but a faculty advisor must serve as the

investigator. Employees at child welfare agencies are not presently eligible to obtain any version of the NSCAW data.

Contact Information:

Data are available to researchers who meet the requirements described in the Access Requirements and Use Restrictions section of this profile through the National Data Archive on Child Abuse and Neglect (NDACAN) at http://www.ndacan.cornell.edu/index.html.

Data Archive information:
National Data Archive on Child Abuse and Neglect
Beebe Hall - FLDC
Cornell University
Ithaca, NY 14853
NDACAN General E-mail: NDACAN@cornell.edu
NDACAN Technical Support: NDACANSupport@cornell.edu

Federal Project Officer:
Mary Bruce Webb, Ph.D.
Administration for Children and Families
370 L'Enfant Promenade, S.W.
Washington, D.C. 20201
Phone: (202) 205-8628
mary.webb@acf.hhs.gov

National Survey of Family Growth (NSFG)

Sponsor: U.S. Department of Health and Human Services (DHHS)/Centers for Disease Control and Prevention (CDC)/National Center for Health Statistics (NCHS)

Description: The National Survey of Family Growth (NSFG) is a periodic survey initiated to provide current information on fertility and infertility, family planning, childbearing, contraceptive practice, and other aspects of maternal and child health and to gauge the effects of these processes on population growth. The NSFG Cycle 6 interviews, conducted in 2002, covered the respondent's pregnancy history, past and current use of contraception, ability to bear children, use of medical services for family planning, infertility, prenatal care, marital history, and associated cohabiting unions. Data on occupation and labor force participation and on a wide range of social, economic, and demographic characteristics are also presented. In addition, Cycle 6 adds detailed questions on HIV risk behaviors and fatherhood and father involvement.

Relevant Policy Issues: Health Disparities, Measures of Well-being for Families/Households, and Identification of Evidence-based Practices and Programs that Improve Family Well-being.

Data Type(s): Survey

Unit of Analysis: Individual

Identification of AI/AN/NA: Race is self-reported to the interviewer (during the face-to-face interview), using the following CAPI instructions:

Which of the groups on Card 2 describe your racial background? Please select one or more groups.
[Interviewers are instructed to enter all that apply. They are to enter all groups that are part of the mixture if the respondent reports a mixture of several races (biracial, mixed, mulatto, etc.).]

- American Indian or Alaska Native (AI/AN)
- Asian
- Native Hawaiian or other Pacific Islander (NH/PI)
- Black or African American
- White

[If respondent selected multiple race groups, interviewer asks this question.]
Which of these groups, that is (Race groups selected) would you say best describes your racial background?

AI/AN/NA Population in Data Set: It is not possible to identify AI/AN/NA persons in the public use file. However, in the restricted Cycle 6 data set (2002), the following cases are available:

AI/AN, Non-Hispanic: 368 (159 men and 209 women)
AI/AN, Hispanic: 579 men and women (most identified themselves as Mexican and South American)
NH/PI: 91 (45 men and 46 women)

Geographic Scope: The geographic scope of the study is national. Detailed geographic identifiers are available on the restricted access contextual data file. These variables include the state code, the county code, census tract, block group, metropolitan status, urban/rural identifiers, and information regarding the land area and population count for the county, tract, and block group.

Analysis can be done for the four major census regions (Northeast, Midwest, South, West) and for metropolitan and nonmetropolitan areas. Estimates cannot be made for individual states or for smaller areas.

Date or Frequency: This is a periodic survey. Previous cycles of data collection for all races include:
Cycle 6: 2002 [7,643 women, 4,928 men (the first time NSFG included a sample of men)]
Cycle 5: 1995 (10,847 women)
Cycle 4: 1988 (8,450 women)
Cycle 3: 1982 (7,969 women)
Cycle 2: 1976 (8,611 women)
Cycle 1: 1973 (9,797 women)

Aggregation: Researchers may wish to pool data from the different cycles of the NSFG to increase the numbers within rare subgroups (such as Native Hawaiians and Other Pacific Islanders). Details on the way race information was collected in earlier cycles of the NSFG are provided below:

Cycle 3 (1982): Question F-47:
"Which of the groups on this card best describes your racial background?"
Alaskan Native or American Indian has 83 cases (out of 7,969 women).
Only one race was coded in Cycle 3.

Cycle 4 (1988): Question F-9:
"Which of the groups on card 30 best describe your racial background?" (Code all that apply)
Alaskan Native or American Indian has 238 cases out of 8,450, allowing multiple mentions. The standard "RACE" recode is shown only as black, white, and other, but the original variable contains more detail.

Cycle 5 (1995): Question IC-3 and 4:
"Which of the groups on Card I-1best describes your racial background?"
First mention= 344 "Alaskan Native or American Indian."
2nd mention = 9 "Alaskan Native or American Indian."
3rd mention = 3 "Alaskan Native or American Indian."

CDC does not recommend using the AI/AN/NA race identifiers on the Cycle 1 (1973) and Cycle 2 (1976) data sets, as this information was collected differently and should not be combined with later cycles.

While there are no CDC-specific guidelines for pooling data, the agency provides the following suggestions for combining the data:

Researchers can append the data from each cycle of interest and use the year of the survey as an independent variable. For obtaining unbiased standard errors for each cycle:

- For the 2002 and 1995 surveys (Cycles 6 and 5) researchers can use SUDAAN or STATA software.
- For 1988 (Cycle 4) there are Balanced Repeated Replicate (BRR) weights available. STATA version 9 does BRR variance estimation. The command and the column locations of the weights can be obtained from CDC.
- For 1982 (Cycle 3): Replicate weights to allow calculation of valid standard errors for complex sampling were not included on public use files. Instead, generalized variance estimates were developed and published for estimated numbers and percentages.

Additionally, researchers should examine the sampling methodology and question text and response options across the different cycles to decide whether pooling data across the different cycles is feasible.

Data Collection Methodology: Data are collected through in-person face-to-face interviews conducted by trained female interviewers. Interviewers use computer assisted personal interviewing (CAPI) to record responses, except for the last section of the questionnaire, which uses audio computer assisted self-interviewing (ACASI) for sensitive questions.

Participation: Optional, with incentives. For Cycle 6, an incentive of $40 was given for completed surveys.

Response Rate: For 2002, the response rate is reported as 80 percent for women and 78 percent for men.

Sampling Methodology: The NSFG employed a multistage national area probability sample design. The target population consisted of women 15-44 years of age in all 50 states and the District of Columbia. The first stage of sampling involved combining all counties in the U.S. to form 2,402 primary sampling units (PSUs). From this, 121 national and Hispanic sample PSUs were selected. Stage 2 divided the sample PSUs into four domains based on estimated key characteristics of the population within a block. From this, a total of 783 segments were selected from the initial sample of 1,414 segments for fieldwork. For stage 3, trained household listers visited each of the sample segments to list housing units on the blocks in the segments. Sampled housing units were drawn from these housing unit lists. The fourth stage consisted of selecting eligible persons from within the sampled households.

National Survey of Family Growth (NSFG)
(continued)

Analysis: The CDC website includes several examples of programs used to create variance estimation for Cycle 6 data of the NSFG. There is not a report that publishes specific design effects for the variables of interest, but the examples included on this website may be very useful to researchers: http://www.cdc.gov/nchs/about/major/nsfg/nsfgvar.htm.

Strengths: Data are collected on key policy issues, including health and family well-being.

Limitations: Since AI/AN/NA persons are not identified in the most recent (2002) public use file, researchers will have to analyze the data through the NCHS Research Data Center. In addition, to increase sample size, researchers may wish to combine the data for 2002 with data from one or more of the previous surveys that collected AI/AN/NA race, conducted in 1995, 1988 and 1982. Researchers are also encouraged to cross-tabulate the data on AI/AN race by Hispanic origin.

In the full sample data, the number of NH/PI is very small.

Access Requirements and Use Restrictions: The AI/AN race categories were included in an "other" category on the public use file due to disclosure risk. To do analysis using these categories separately, a researcher may use the Research Data Center (RDC) at NCHS, which is a physical space located within the NCHS facilities in Hyattsville, Maryland, where researchers are allowed access to NCHS restricted data files not released to the public. These data files do not contain direct identifiers such as name or social security number, but may contain identifiers for small geographic units such as block or census tract.

There are 3 ways to access data through the RDC once a project has been approved:
(1) “remote access,” in which the user submits a SAS program electronically and the output is screened and returned electronically;
(2) by physically going to the RDC at NCHS, and doing the research there, using any software the researcher wants to use; and
(3) staff-assisted access, in which the researcher submits a program to an RDC staff member who submits and screens it and returns the output.
There are fees for each type of access.

Contact Information:
National Survey of Family Growth Staff
Division of Vital Statistics
National Center for Health Statistics
Centers for Disease Control and Prevention
3311 Toledo Road, Floor 7
Hyattsville, Maryland 20782
(301) 458-4222
nsfg@cdc.gov

Research Data Center: (301) 458-4277 or e-mail rdca@cdc.gov

National Survey of Veterans (NSV)

Sponsor: U.S. Department of Veterans Affairs

Description: The 2001 National Survey of Veterans (NSV) is the fifth in a series of comprehensive nationwide surveys designed to help the Department of Veterans Affairs (VA) plan its future programs and services for veterans. The information gathered through these surveys will help VA to identify the needs of veterans and then allocate resources in ways that will ensure these needs can be met. It also provides a snapshot profile of the veteran population. Data collected through the NSV enables VA to: (1) follow changing trends in the veteran population; (2) compare characteristics of veterans who use VA services with those of veterans who do not; (3) study VA's role in the delivery of all benefits that veterans receive; and (4) update information about veterans to help the VA develop its policies.

Relevant Policy Issues: Eligibility and Use of Veterans Administration Health Facilities and Eligibility and Use of Other Veterans Administration Benefits.

Data Type(s): Survey

Unit of Analysis: Individual

Identification of AI/AN/NA: The instructions for reporting race were as follows:
I'm going to read a list of racial categories. Please select one or more to describe your race. Are you…

- White
- Black or African American
- American Indian or Alaska Native (AI/AN)
- Asian
- Native Hawaiian (NH)
- Other Pacific Islander (OPI)
- Hispanic/Mexican
- Other

Respondents could choose up to 8 response categories for this item.

AI/AN/NA Population in Data Set: The counts reported below represent the number of respondents who mentioned the listed race as their race either with or without mentioning any other race(s).

TOTAL: 20,048
AI/AN: 897
NH: 34
OPI: 48

National Survey of Veterans (NSV)
(continued)

AI/AN/NA Subpopulations: AI/AN/NA subpopulations identified are:

- Native Hawaiian alone
- Pacific Islander alone

Geographic Scope: The geographical scope is national. No additional geographic areas were identified in the data.

Date or Frequency: The National Survey of Veterans was first conducted in the late 1970s, again in 1987, again in 1993, and most recently in 2001. The content of the questionnaires changed between data collection efforts.

Data Collection Methodology: The 2001 NSV was a computer-assisted telephone interview (CATI).

Participation: Optional, without incentives

Response Rate: There were two types of samples: random-digit dialing (RDD) and list. Overall response rate was 51.6 percent for the RDD sample. The overall list sample response rate was not presented in the methodology report.

Sampling Methodology: The sample design for the 2001 NSV was a dual frame design consisting of an RDD sample and a list sample. The list sample design used the Veterans Health Administration Healthcare enrollment file and the Veterans Benefits Administration Compensation and Pension (C&P) file to construct the sampling frame.

Analysis: Appendix C of the final report includes a discussion of standard errors for estimates. Information on the NSV is available online at: http://www.virec.research.va.gov/DataSourcesName/NationalSurveyVeterans/2001NationalSurveyofVeterans.htm

Authorization: The NSV is conducted under the general authorization of U.S. Code Title 38, Section 527. This section authorizes the VA Secretary to gather data for the purposes of planning and evaluating VA programs.

Strengths: Data are collected on key policy issues. There are multiple years of data available. The NSV is one of the few national data sets that covers veterans' issues. Also, comprehensive documentation of the study is available online.

Limitations: There are only a small number of NH/PI respondents. Although the NSV has been conducted in previous years, there are significant differences in the survey content across the administrations. Aggregation of the data to increase the number of NH/PI represented in the database would require a skilled statistician.

National Survey of Veterans (NSV)
(continued)

Access Requirements and Use Restrictions:	Current data are available to the public at no cost; this could change if demand is high. Data from these earlier efforts may be available depending on demand.
Contact Information:	In order to obtain this file, researchers should contact the Office of Policy, Planning and Prepardness Office, Office of Policy, Department of Veterans Affairs, and the 2001 National Survey of Veterans Project Officer, Susan Krumhaus at (202) 273-5108, or Wayne Johnson at (202) 273-8972, and briefly explain how you would like to use these data.

National Survey on Drug Use and Health (NSDUH)

Sponsor: U.S. Department of Health and Human Services (DHHS)/Substance Abuse and Mental Health Services Administration (SAMHSA)/Office of Applied Studies (OAS)

Description: The National Survey on Drug Use and Health (NSDUH), formerly called the National Household Survey on Drug Abuse or NHSDA, is designed to produce drug and alcohol use incidence and prevalence estimates and report the consequences and patterns of use and abuse in the general U.S. civilian population aged 12 and older. Questions include age at first use, as well as lifetime, annual, and past-month usage for many drugs. The survey also covers substance abuse treatment history and perceived need for treatment, and includes questions from the Diagnostic and Statistical Manual (DSM) of Mental Disorders that allow diagnostic criteria to be applied. Respondents are also asked about personal and family income sources and amounts, health care access and coverage, illegal activities and arrest record, problems resulting from the use of drugs, perceptions of risks, and needle-sharing. Demographic data include gender, race, age, ethnicity, educational level, job status, income level, veteran status, household composition, and population density.

Relevant Policy Issues: Key Health Disparities.

Data Type(s): Survey

Unit of Analysis: Individual

Identification of AI/AN/NA: Which of these groups describes you? (The interviewer gives the respondent a handcard with race categories and instructs respondent to provide one or more races.)

- White
- Black/African American
- American Indian or Alaska Native (AI/AN) (American Indian includes North American, Central American, and South American Indians)
- Native Hawaiian (NH)
- Other Pacific Islander (OPI)
- Asian (for example, Asian Indian, Chinese, Filipino, Japanese, Korean, and Vietnamese)
- Other

In the public use data, NH and OPI are combined into a single category.

AI/AN/NA Population in Data Set: The achieved sample for the 2004 NSDUH was 67,760 persons. The public use file contains 55,602 records due to a subsampling step used in the disclosure protection procedures.

From the public use 2004 data:
AI/AN (coded as Non-Hispanic Native American/Alaska Native): 784
NH/PI (coded as Non-Hispanic NH/PI): 218

Geographic Scope: The geographic scope of the study is national. Geographic indicators available on the public-use file include Metropolitan Statistical Area (MSA) with 1 million or more people, MSA with less than 1 million people, and not in MSA.

Date or Frequency: This is an annual study that has been collected from 1971 to the present. The most recent year of available data is 2004. Data from the previous year's collection are typically released in September of the following year. Data from 2005 are expected to be released in September 2006.

Data Collection Methodology: The NSDUH is administered in-person by field interviewers at respondents' residences.

Participation: Optional, with incentives

Response Rate: The study yielded a weighted screening response rate of 91 percent and a weighted interview response rate for the computer assisted personal interview (CAPI) of 77 percent.

Sampling Methodology: The NSDUH uses a multistage area probability sample for each of the 50 states and the District of Columbia. The 2004 sample design is a continuation of the coordinated five-year sample design that increases the precision of estimates in year-to-year trend analysis. The sample is stratified on multiple levels, beginning with states. The second level of stratification divides states into field interviewer (FI) regions. For the first stage of sampling, each FI region is partitioned into small geographic areas composed of adjacent census blocks (segments). Systematic sampling is then used to select the allocated sample of addresses from each segment. The sample design includes approximately equal numbers of persons in the following age groups: 12-17, 18-25, and 26 and older.

Analysis: There are three different analysis weights available with the 2004 NSDUH data. One weight is used when analyzing variables asked of all respondents. The others are used when analyzing data asked only of a subgroup of respondents (resulting from a complex split-sample design). Detailed instructions for applying these weights for analysis can be found at http://www.icpsr.umich.edu/SDA/SAMHDA/04373-0001/CODEBOOK/4373.htm.

Strengths: Data are collected on a key policy issue, health. There are multiple years of data available. The 2004 NSDUH is specifically designed to facilitate precise trend analysis using prior years of the survey data.

National Survey on Drug Use and Health (NSDUH)
(continued)

Limitations: The NSDUH does not collect data from persons who are homeless who do not stay at shelters, active duty military personnel, and persons housed in jails or hospitals.

Access Requirements and Use Restrictions: Data sets are available to the public at no cost.

Contact Information: Data and documentation can be downloaded at:
http://webapp.icpsr.umich.edu/cocoon/SAMHDA-STUDY/04373.xml

Data Archive Information:
Substance Abuse and Mental Health Data Archive (SAMHDA)
SAMHDA Helpline: (888) 741-7242
Local: (734) 615-9524
Fax: (734) 647-8200
e-mail: samhda-support@icpsr.umich.edu

SAMHDA/ICPSR
The University of Michigan
P.O. Box 1248
Ann Arbor, MI 48106-1248
U.S.A.

General Inquiries should be addressed to:
Joe Gustin
Assistant Project Officer
DHHS/SAMHSA/OAS
1 Choke Cherry Road, Room 7-1020
Rockville, MD 20857
e-mail: Joe.Gustin@samhsa.hhs.gov
http://www.oas.samhsa.gov/nsduh.htm

National Vital Statistics System: Linked Birth-Infant Death (NVSS-I)

Sponsor: U.S. Department of Health and Human Services (DHHS)/Centers for Disease Control and Prevention (CDC)/National Center for Health Statistics (NCHS)

Description: The National Vital Statistics System Linked Birth-Infant Death (NVSS-I) research data set is comprised of linked birth and death certificates for infants born in the United States, Puerto Rico, the Virgin Islands, and Guam who died before reaching 1 year of age. In this data set, information from the death certificate is linked with information from the birth certificate for each infant. The purpose of this linkage is to use the many additional variables available from the birth certificate in infant mortality analysis. The birth certificate is the primary source of demographic information, such as age, race, and Hispanic origin of the parents; maternal education; live birth order; and mother's marital status; and of maternal and infant health information, such as birthweight, period of gestation, plurality, prenatal care usage, and maternal smoking, etc. Analysis of this information can provide insight into the major factors influencing infant mortality in the United States. This system of linked records was established in 1983.

Relevant Policy Issues: Measurement of Health Status, Health Disparities.

Data Type(s): Registry

Unit of Analysis: Individual

Identification of AI/AN/NA: The majority of the demographic information is obtained from the birth certificate. Below is an example of the race question as it appears on the U.S. Standard Certificate of Live Birth:

MOTHER'S RACE (Check one or more races to indicate what the mother considers herself)

- White
- Black or African American
- American Indian or Alaska Native (AI/AN) and (Name of the enrolled or principal tribe)
- Asian Indian
- Chinese
- Filipino
- Japanese
- Korean
- Vietnamese
- Other Asian (Specify)
- Native Hawaiian (NH)
- Guamanian or Chamorro
- Samoan

- Other Pacific Islander (OPI) (Specify)
- Other (Specify)

Beginning in 2003, the number of births for any of the Asian /Pacific Islander subgroups is no longer available. Please see the NVSS report entitled *Births: Final Data for 2003* for an explanation:
http://www.cdc.gov/nchs/data/nvsr/nvsr54/nvsr54_02.pdf

AI/AN/NA Population in Data Set:

From the 2003 linked file (*National Vital Statistics Report, Volume 54, Number 16*):
Race of Mother is American Indian (AI)
Births: 43,054 (N=4,090,007)
Infant deaths: 376* (N=27,995)
Neonatal deaths: 196 (N=18,935)
Post-neonatal deaths: 180 (N=9,060)
*Infant deaths are weighted, so numbers may not exactly add to totals due to rounding.

From the 2002 linked file (*National Vital Statistics Report, Volume 53, Number 10*):
Race of Mother is Hawaiian (NH)
Births: 6,772 (N=4,021,825)
Infant deaths: 65* (N=27,970)
Neonatal deaths: 38 (N=18,791)
Post-neonatal deaths: 27 (N=9,179)
*Infant deaths are weighted, so numbers may not exactly add to totals due to rounding.

AI/AN/NA Subpopulations:

Based on available reports, identification for Native Hawaiians is possible for some years of data linkage (e.g., 2002).

Geographic Scope:

Geographic scope of the data is national. Place of birth and place of death are classified by state and county. In residence classification of the birth, all births are allocated to the usual place of residence of the mother as reported on the birth certificate and are classified by state, county, and city. In residence classification of the death, all deaths are allocated to the usual place of residence of the decedent as reported on the death certificate and are classified by state, county, and city. Counties and cities of 250,000 persons or more are identified in the linked data set. Geographic classification for the linked data set is based on the 1980 census enumeration.

Date or Frequency:

Linked files are available for the data years 1983-91 and 1995-2002. Linked file data were not produced for the 1992-94 data years. Future data years will be available annually.

Data Collection Methodology:

Vital statistics are provided through state-operated registration systems. Administrative records pertaining to death certificates are completed by

physicians, coroners, medical examiners, and funeral directors. Administrative records pertaining to birth certificates are completed by physicians and midwives. These records are filed with state vital statistics offices and selected statistical information is forwarded to NCHS to be merged into a national statistical file.

Participation: Mandatory

Strengths: Data sets contain a moderate number of AI/AN/NA respondents. There are multiple years of data available.

Limitations: There is limited documentation available for this study, so it is difficult to know which subpopulations of key interest can be examined separately using the data files. For example, in the 2002 report, Hawaiians are reported separately, but they are not reported separately in the 2003 report.

Access Requirements and Use Restrictions: Linked Birth and Infant Death public use data are available for the years 1983-91 and 1995-98 on CD-ROM. The Linked Birth and Infant Death CD-ROM can be purchased through the National Technical Information Service (NTIS) and/or the Government Printing Office (GPO). As prices vary, contact NTIS or GPO for current pricing.

Contact Information: Reproductive Statistics Branch
Division of Vital Statistics
National Center for Health Statistics
Centers for Disease Control and Prevention
3311 Toledo Road, Room 7417
Hyattsville, Maryland 20782
(301) 458-4356

National Vital Statistics System: Mortality (NVSS-M)

Sponsor: U.S. Department of Health and Human Services (DHHS)/Centers for Disease Control and Prevention (CDC)/Coordinating Center for Health Information and Service (CCHIS)/National Center for Health Statistics (NCHS)

Description: The National Vital Statistics System Mortality (NVSS-M) data set is generated from death certificate information collected through the National Vital Statistics System, an inter-governmental collaboration between NCHS and the 50 states, two cities, and five territories. The NVSS-M data serve as the primary source of information on demographic, geographic, and cause-of-death information among persons dying in a given year. Data are available on an annual basis. Variables include the following: year, month, and day of week of death; place of death; residence of decedent (state, county, city, population size, standard metropolitan statistical area, metropolitan and nonmetropolitan counties); state and county of occurrence; demographic information on decedent (e.g., age at death, education, Hispanic origin, marital status, race, sex, state of birth); underlying cause of death; and multiple causes of death.

Relevant Policy Issues: Measurement of Health Status, Key Health Disparities.

Data Type(s): Registry

Unit of Analysis: Individual

Identification of AI/AN/NA: Races available on public use data sets:

- White
- Black
- American Indian (includes Aleuts and Eskimos)
- Chinese
- Japanese
- Hawaiian (includes Part-Hawaiian)
- Filipino
- Asian Indian
- Korean
- Samoan
- Vietnamese
- Guamanian
- Other Asian or Pacific Islander*
- Combined other Asian or Pacific Islander**

* Other Asian or Pacific Islander includes any Asian or Pacific Islander (API) group that is not included and does not easily fall into one of the API categories listed above.

** The "Combined Other Asian or Pacific Islander" category are death records in which more than one API category was listed under "race" on the death certificate (multiple race deaths). This category was introduced with the 2003 data. Most of the deaths in this category come from California and Hawaii (98.3 percent of the category).

AI/AN/NA Population in Data Set: The numbers below represent the 2003 NVSS-M public use data set:

Total number of records: 2,452,154
White: 2,106,697
Black: 291,706
American Indian: 13,160
Chinese: 8,831
Japanese: 5,920
Hawaiian: 594
Filipino: 7,557
Asian Indian: 2,542
Korean: 2,548
Samoan: 404
Vietnamese: 2,024
Guamanian: 159
Other Asian or Pacific Islander: 8,438
Combined other Asian or Pacific Islander: 1,574

AI/AN/NA Subpopulations: AI/AN/NA subpopulations identified are:

- Native Hawaiian alone (in detailed data only)
- Samoan (in detailed data only)
- Guamanian (in detailed data only)

Geographic Scope: The geographic scope of the data includes national, region, state and counties with population of 100,000 or more.

Date or Frequency: Public use data sets from 1968 through 2003 are available.

Data Collection Methodology: Data from all death certificates filed in the United States are compiled into an annual file, except for 1972 when only a 50 percent sample was compiled.

Participation: Mandatory

Strengths: Data sets may contain a large number of AI/AN/NA respondents. There are multiple years of data available.

Limitations: The primary weaknesses of these data sources are the quality of the race/ethnicity data. Because data are extracted from death certificates, the race/ethnicity category is not self-reported and is often completed by a funeral director based on

information received from a family member/proxy on the race/ethnicity of the deceased individual. According to the NCHS, these data are particularly poor for the American Indian/Alaska Native category, as data quality checks of the racial/ethnic distribution of the deceased in this category are lower than the distribution represented in Census estimates.

The Indian Health Service (IHS) conducted a study on racial misclassification of mortality data and produced adjustments to apply to the rates to compensate for the misreporting of AI/AN race on the state death certificates. This study found that misidentification varies substantially among states and the IHS service areas. IHS now adjusts the NVSS-M data to correct this misidentification and presents both adjusted and unadjusted data in two major IHS publications entitled: *Regional Differences in Indian Health* and *Trends in Indian Health*. (The reports are currently being updated to include the latest adjustment factors.)

Access Requirements and Use Restrictions:

Public use data sets are available at no cost. Potential users are advised to contact the Mortality Statistics Branch at (301) 458-4666.

Data on injury deaths for the race categories "American Indian/Alaskan Native" and "Asian/Pacific Islander" are available from CDC's web-based Injury Statistics Query and Reporting System (WISQARS) at: http://www.cdc.gov/ncipc/wisqars/

Contact Information:

General Contact for Mortality statistics data:
Robert N. Anderson, Ph.D.
Branch Chief, Mortality Statistics Branch
Division of Vital Statistics
National Center for Health Statistics
Centers for Disease Control and Prevention
3311 Toledo Rd., Room 7318
Hyattsville, Maryland 20782
(301) 458-4073

For information on accessing mortality files, contact: Ken Kochanek (301) 458-4319

National Vital Statistics System: Natality (NVSS-N)

Sponsor: U.S. Department of Health and Human Services (DHHS)/Centers for Disease Control and Prevention (CDC)/National Center for Health Statistics (NCHS)

Description: The National Vital Statistics System Natality (NVSS-N) public-use data file comprises records of all documented births occurring within the United States. Data from all birth certificates filed in each state are included in this file. Specifically, these data cover the following information: residence of mother (e.g., population size of residence community, standard metropolitan statistical area, metropolitan and nonmetropolitan counties); demographic information about parents (e.g., race of parents, age of parents, education of mother and father, pregnancy/childbearing history of mother, marital status of mother); information on the infant (e.g., race, sex, Apgar scores at 5 minutes after birth, total-birth order); and information on the birth [e.g., place of birth, place of delivery, birth date (month/day), birth weight (in grams), gestation period, and prenatal care].

Relevant Policy Issues: Measurement of Health Status, Key Health Disparities, and Factors Contributing to Measured Health Disparities.

Data Type(s): Registry

Unit of Analysis: Individual

Identification of AI/AN/NA: Data are coded into the following categories:

- White
- Black
- American Indian/Alaska Native (AI/AN)
- Asian/Pacific Islander (microdata early than 2003 are subdivided into: Chinese, Japanese, Filipino, Hawaiian, Other API)

There are no detailed breakdowns for any of the Asian/Pacific Islander subgroups after 2002. Please see the NVSS Report "Births: Final Data for 2003" for an explanation: http://www.cdc.gov/nchs/data/nvsr/nvsr54/nvsr54_02.pdf

Beginning 2003, data are available for selected states for multiple race reporting for the mother and the father. As of the 2004 data year, 15 states reported multiple race responses to NCHS. Because most states do not report multiple race, it is necessary to “bridge” the multiple race responses to single race, following a special algorithm developed by NCHS in cooperation with the Census Bureau and with support from the National Cancer Institute. A large proportion of births to American Indian and Hawaiian women in particular are to women reporting more than one race. Detailed verbatim and checkbox entries for multiple race persons are available on the natality files by request.

National Vital Statistics System: Natality (NVSS-N)
(continued)

AI/AN/NA Population in Data Set:	Total number of births registered in the United States in 2004: 4,112,052 AI/AN registered births in 2004: 43,927
Geographic Scope:	The geographic scope of the data is national. Geographic areas identified are state, county, city (if more than 100,000 population), standard metropolitan statistical area (SMSA), and metropolitan/nonmetropolitan counties. Additional analyses are possible by state, county, city (if more than 100,000 population), standard metropolitan statistical area (SMSA), metropolitan/nonmetropolitan counties.
Date or Frequency:	Data from all birth certificates are compiled into an annual file, except for the files for 1951-54, 1956-66, and 1968-71, when only a 50 percent sample was compiled, and 1967 when a 20- to 50-percent sample was compiled. Data for 1972-84 are based on 100 percent of births for selected states and a 50 percent sample for all other states.
Data Collection Methodology:	Data from all birth certificates are compiled into an annual file, except for the files 1972, 1981 and 1982 when only a 50 percent sample was compiled.
Participation:	Mandatory for NCHS to compile a national data set; state participation is based on the Vital Statistics Cooperative Program of the National Vital Statistics System.
Strengths:	Data sets contain a large number of AI/AN respondents. Data are collected on key health issues. There are multiple years of data available.
Limitations:	Of the Pacific Islander population groups, only Native Hawaiians as either a single- or multiple-race category can be separately identified in the microdata.
Access Requirements and Use Restrictions:	The data set is available to the public on CD-ROM at no charge from births@cdc.gov or by contacting the office listed below.
Contact Information:	Information on the Public Use Files and instructions for obtaining files can be located at http://www.cdc.gov/nchs/products/elec_prods/subject/natality.htm, or by contacting births@cdc.gov. For custom data requests, contact: Reproductive Statistics Branch Division of Vital Statistics National Center for Health Statistics Centers for Disease Control and Prevention 3311 Toledo Road, Floor 7318 Hyattsville, Maryland 20782 (301) 458-4111 births@cdc.gov

Panel Study of Income Dynamics (PSID)

Sponsor: The Panel Study of Income Dynamics (PSID)'s original funding agency was the Office of Economic Opportunity of the United States Department of Commerce. The study's major funding source is now the National Science Foundation. Substantial additional funding has been provided by: the National Institute on Aging, the National Institute of Child Health and Human Development and the Office of the Assistant Secretary for Planning and Evaluation of the United States Department of Health and Human Services; the Economic Research Service of the United States Department of Agriculture; the United States Department of Housing and Urban Development; the United States Department of Labor; and the Center on Philanthropy at the Indiana University-Purdue University.

Description: The PSID, begun in 1968, is a longitudinal study of a representative sample of U.S. individuals (men, women, and children) and the family units in which they reside. Its emphasis is on economics, but it also includes sociological and psychological measures.

Relevant Policy Issues: Income Status, Unemployment Rates, Economic Assistance Program Participation Rates, Economic Opportunity, Measures of Well-being for Families/households, Measures of Well-being for Children, Measures of Well-being for Elders, and Housing Ownership.

Data Type(s): Survey

Unit of Analysis: Individual

Identification of AI/AN/NA: Instructions for reporting race are as follows: “In order to get an idea of the different races and ethnic groups that participate in the study, I would like to ask you about your background. Are you:

- White
- Black
- Native American (NA)
- Asian
- Pacific Islander
- Another race

Up to four choices were recorded.

AI/AN/NA Population in Data Set: There are 7,822 families in the full 2003 data set. The unweighted count for NA male heads of households of these families is 136 (41 are NA alone and 95 are NA and other races). The unweighted count for wives of heads of households in the 2003 wave of the PSID is 39 (24 are NA alone and 15 are NA and other races).

(Pacific Islanders are combined with Asians into a single group, so no separate count of this group is possible.)

Panel Study of Income Dynamics (PSID)
(continued)

Geographic Scope: The geographic scope of the study is national. The public release files contain geographic information such as region, state of residence, size of largest city in the county of residence, and the Beale rural-urban code. The Beale rural-urban code includes the following categories:

- Fringe counties of metropolitan areas of 1 million population or more
- Counties in metropolitan areas of 250,000 to 1 million population
- Counties in metropolitan areas of less than 250,000 population
- Urban population of 20,000 or more, adjacent to metropolitan area
- Urban population of 20,000 or more, not adjacent to a metropolitan area
- Urban population of less than 20,000, adjacent to a metropolitan area
- Urban population of less than 20,000, not adjacent to a metropolitan area
- Completely rural, adjacent to a metropolitan area
- Completely rural, not adjacent to a metropolitan area

The data allow geographic analysis at all these levels.

Date or Frequency: Between 1968 and 1997, PSID data were collected every year. Starting in 1999, the PSID collected data biennially (i.e., every other year). All waves of data 1968-2003 are available on the website. The 2005 data will be released by December 31, 2006. The next wave of the PSID will be conducted in 2007.

Data Collection Methodology: The PSID was collected in face-to-face interviews using paper and pencil questionnaires between 1968 and 1972. Thereafter, the majority of interviews were conducted over the telephone. In 1993, the PSID introduced the use of computer assisted telephone interviewing. In the 1999 wave, 97.5 percent of the interviews were conducted over the phone, and all interviews were conducted using computer-based instruments.

Participation: Optional, with incentives

Response Rate: Since 1969, annual response rates have ranged between 96.9 and 98.5 percent.

Sampling Methodology: The initial sample for the PSID consisted of two independent samples: a cross-sectional, national sample (based on stratified multistage selection of the civilian noninstitutional population of the U.S.) and a national sample of low-income families. Both samples are probability samples. However, when the two samples are combined the result is a sample with unequal selection probabilities, and as a result compensatory weighting is needed in estimation.

Strengths: Data are collected on key policy issues, including economic status and child well-being. There are multiple years of data available. Documentation of the content and implementation of the PSID is comprehensive and available on-line.

Limitations: There are a very small number of AI/AN/NA respondents. Asians and Pacific Islanders are collapsed into a single result category in all PSID data sets.

Panel Study of Income Dynamics (PSID)
(continued)

Access Requirements and Use Restrictions:	Data set is available to the public at no cost.
Contact Information:	The data set is available from the following website: http://simba.isr.umich.edu/. For general assistance, contact: PSID Staff The Panel Study of Income Dynamics Institute for Social Research PO Box 1248 Ann Arbor, MI 48106-1248 psidhelp@isr.umich.edu

Pediatric Nutrition Surveillance System (PedNSS)

Sponsor: U.S. Department of Health and Human Services (DHHS)/Centers for Disease Control and Prevention (CDC)/National Center for Chronic Disease Prevention and Health Promotion/Division of Nutrition and Physical Activity/Maternal and Child Nutrition Branch

Description: The Pediatric Nutrition Surveillance System (PedNSS) is a child-based public health surveillance system that monitors the nutritional status of low-income children in federally funded maternal and child health programs. Data on birthweight, breastfeeding, anemia, short stature, underweight, and overweight are collected for children who attend public health clinics for routine care, nutrition education, and supplemental food. Data are collected at the clinic level then aggregated at the state level and submitted to CDC for analysis. Online national PedNSS data are available as published tables. State-level online tables are also available for both California and West Virginia in addition to online national-level tables.

Relevant Policy Issues: Measurement of Health Status, Disease-specific Measurements, Health Disparities, and Measures of Well-being for Children.

Data Type(s): Program reporting data. Data are collected from children enrolled in federally funded programs that serve low-income children, including the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) and non-WIC programs that include the Early and Periodic Screening, Diagnosis, and Treatment (EPSDT) Program and the Title V Maternal and Child Health Program.

Unit of Analysis: Individual (unique child records)

Identification of AI/AN/NA: Data include the following racial/ethnic categories:

- White, not Hispanic
- Black, not Hispanic
- Hispanic
- American Indian/Alaska Native (AI/AN)
- Asian/Pacific Islander
- All Other

AI/AN/NA Population in Data Set: For 2004, the total number of individual children was 6,822,769. Of these, 77,915 were identified as AI/AN.

AI/AN/NA Subpopulations: State data are broken out by race and include some tribe identifiers: Cheyenne River Sioux (SD), InterTribal Council of Arizona, Rosebud Sioux Tribe (SD), Chickasaw Nation (OK), Wichita-Caddo-Delaware (OK), Navajo Nation (AZ), Standing Rock Sioux Tribe (ND).

Pediatric Nutrition Surveillance System (PedNSS)
(continued)

Geographic Scope:	PedNSS is a national surveillance system. PedNSS Surveillance data are reported from contributors (defined as a state, U.S. territory, or tribal government). In 2004, a total of 48 contributors, including 40 states, the District of Columbia, Puerto Rico, and 7 tribal governments, participated in PedNSS. PedNSS is a voluntary surveillance system. Online geographic analysis is possible through review of published tables of national data and for two states at the websites listed below: California has its data available at the following website: http://www.dhs.ca.gov/pcfh/cms/onlinearchive/pdf/chdp/informationnotices/2003/chdpin03q/contents.htm. West Virginia provides state-specific data from 1996 - present http://www.wvdhhr.org/ons/surveillance.asp.
Date or Frequency:	Trend data tables present data from 1995 – 2004.
Data Collection Methodology:	Federally funded health clinics serving low-income children participate on a voluntary basis and report data to state-level agencies, which in turn submit data to the CDC. These data are combined for annual reporting.
Participation:	Optional, without incentives
Strengths:	Registry contains a large number of AI/AN/NA respondents. Data are collected on key policy issues, including health and child welfare. There are multiple years of data available.
Limitations:	Pacific Islanders are not separated from Asians. Not all states, or federally funded clinics within states, participate in this surveillance system; therefore, data are not representative of all children served by programs such as the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC); Early and Periodic Screening, Diagnostic, and Treatment Services (EPSDT); and other Maternal and Child Health Bureau programs.
Access Requirements and Use Restrictions:	National data set is not available to the public, but published tables and reports are available.
Contact Information:	National PedNSS data tables can be accessed through the following website: http://www.cdc.gov/pednss. State-level data for California and West Virginia are available. See above for information on Internet locations of these data.

Pregnancy Nutrition Surveillance System (PNSS)

Sponsor:	U.S. Department of Health and Human Services (DHHS)/Centers for Disease Control and Prevention (CDC)
Description:	PNSS is a program-based public health surveillance system that monitors risk factors associated with infant mortality and poor birth outcomes among low-income pregnant women who participate in federally funded public health programs including Special Supplemental Nutrition Program for Women, Infants, and Children (WIC), and Title V, the Maternal and Child Health Program (MCH). Data include indicators of maternal health and maternal health behavior including pre-pregnancy weight status, parity, and diabetes. National PNSS data are available as published tables. States have the option of making the data publicly available. North Carolina, California, and West Virginia make state data available for download through the Internet.
Relevant Policy Issues:	Measurement of Health Status, Disease-specific Measurements, and Health Disparities.
Data Type(s):	Program reporting data
Unit of Analysis:	Individual
Identification of AI/AN/NA:	The racial/ethnic categories used in this data set include the following: • White, not Hispanic • Black, not Hispanic • Hispanic • American Indian/Alaska Native (AI/AN) • Asian/Pacific Islander • All Other
AI/AN/NA Population in Data Set:	Of the 856,123 total records in 2004, 11,686 were identified as AI/AN.
AI/AN/NA Subpopulations:	State data are broken out by race and include some tribe identifiers: Cheyenne River Sioux (SD), InterTribal Council of Arizona, Rosebud Sioux Tribe (SD), Chickasaw Nation (OK), Navajo Nation (AZ), Standing Rock Sioux Tribe (ND).
Geographic Scope:	PNSS is a national surveillance system. Other geographic identifiers include states (all except for Washington, Alaska, New Mexico, Oklahoma, Texas, Virginia, North Carolina, Connecticut, Massachusetts, Mississippi, Arizona) and selected tribes (see above).

Pregnancy Nutrition Surveillance System (PNSS)
(continued)

	Geographic analysis is possible at the national level through review of the published tables. Additionally, California has its data available at the following website: http://www.dhs.ca.gov/pcfh/cms/onlinearchive/pdf/chdp/informationnotices/2003/chdpin03q/contents.htm. North Carolina provides state-specific data from 1997 - present http://www.nutritionnc.com/nutrsurv.htm. West Virginia provides state-specific data from 1998 - present http://www.wvdhhr.org/ons/surveillance.asp.
Date or Frequency:	Trend tables present data from 1994 – 2003.
Data Collection Methodology:	Federally funded health clinics serving pregnant women participate on a voluntary basis and report data to state-level agencies, which in turn submit the data to the CDC. These data are combined for annual reporting.
Participation:	Optional, without incentives
Strengths:	Registry contains a large number of AI/AN/NA respondents. Data are collected on key policy issues, including health, particularly maternal risk factors associated with infant mortality and peer birth outcomes. There are multiple years of data available.
Limitations:	Pacific Islanders are not separated from Asians. Not all states, or federally funded clinics within states, participate in this surveillance system; therefore, data are not representative of pregnant women served by programs such as the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC); Early and Periodic Screening, Diagnostic, and Treatment Services (EPSDT); and other Maternal and Child Health Bureau programs. The racial/ethnic distribution of individuals served by contributing clinics is presented by state and for some tribes; however, there is no such geographic breakdown for any of the health indicators collected in the data source in the published tables.
Access Requirements and Use Restrictions:	National data set is not available to the public, but published tables and reports are available.
Contact Information:	Data tables can be accessed through the following website: http://www.cdc.gov/pednss/pnss_tables/index.htm State-level data are available from North Carolina, California, and West Virginia. See above for Internet locations of these data.

Pregnancy Risk Assessment Monitoring System (PRAMS)

Sponsor: U.S. Department of Health and Human Services (DHHS)/Centers for Disease Control and Prevention (CDC)

Description: The Pregnancy Risk Assessment Monitoring System (PRAMS) was initiated in 1987 to monitor maternal experiences and attitudes before, during, and shortly after pregnancy to better understand adverse outcomes of mothers and infants. PRAMS collects the following data: state, most core birth certificate variables (not included are birth certificate number; specific date of the month in the infant's date of birth, mother's date of birth, and mother's date of last menses; county of residence; and hospital of birth). On a monthly basis, a sample of women (approximately 1,300-3,400 women per state) who are state residents and have delivered a live-born infant during the preceding 2-4 months are randomly selected (with an oversample of women at higher risk for adverse pregnancy outcomes) from a file of birth certificate records and mailed a questionnaire. Core questions in this instrument include:

- Attitudes and feelings about the most recent pregnancy,
- Content and source of prenatal care,
- Maternal alcohol and tobacco consumption,
- Physical abuse before and during pregnancy,
- Pregnancy-related morbidity,
- Infant health care,
- Contraceptive use, and
- Mother's knowledge of pregnancy-related health issues, such as adverse effects of tobacco and alcohol; benefits of folic acid; and risks of HIV.

Thirty–seven states, New York City, and Yankton Sioux Tribe of South Dakota currently participate in PRAMS.

Relevant Policy Issues: Measurement of Health Status, Disease-specific Measurements, Key Health Disparities, and Factors Contributing to Well-being Disparities of Children.

Data Type(s): Survey

Unit of Analysis: Individual

Identification of AI/AN/NA: Race is identified in the data set in the following categories:

- White
- Black
- American Indian or Alaska Native (AI/AN)
- Asian/Pacific Islander (subdivided into: Chinese, Japanese, Filipino, Hawaiian, Other API)

Pregnancy Risk Assessment Monitoring System (PRAMS)
(continued)

AI/AN/NA Population in Data Set:	An overall total number of AI/AN/NA respondents in the data sets was not available. However, an analysis of 8 PRAMS participating states (see link below) indicated that several states have a sample of 30 or greater American Indian/Alaska Native respondents. This suggests that for analyses aggregated to the national level, there is sufficient sample size for analyses by AI/AN. http://www.cdc.gov/mmwr/preview/mmwrhtml/ss5304a1.htm.
Geographic Scope:	Thirty–seven states, New York City, and the Yankton Sioux Tribe of South Dakota currently participate in PRAMS. Six other states previously participated. The currently participating states are Alabama, Alaska, Arkansas, Colorado, Delaware, Florida, Georgia, Hawaii, Illinois, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Nebraska, New Jersey, New Mexico, New York, North Carolina, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin, and Wyoming.
Date or Frequency:	PRAMS data are available in annual files by individual participating state. The availability of years 1988 to 2004 varies. The 2004 data are the most recently available data set. Data availability by state and year can be reviewed at http://www.cdc.gov/prams/index.htm.
Data Collection Methodology:	PRAMS utilizes two sequential modes of data collection; a mailed questionnaire survey with multiple follow-up attempts to encourage response was followed by a telephone survey.
Participation:	Optional, without incentives.
Response Rate:	The following reference notes that the median response rate across participating states was 76 percent among surveyed mothers: *The Pregnancy Risk Assessment Monitoring System (PRAMS): Current Methods and Evaluation of 2001 Response Rates.* Public Health Rep. 2006 Jan-Feb;121(1):74-83.
Sampling Methodology:	Each participating state draws a stratified systematic sample of 100 to 250 new mothers every month from a frame of eligible birth certificates (mother recently gave live birth), with most states oversampling low birth weights.
Analysis:	The PRAMS data set includes weights to adjust for non-response bias and to help generate accurate standard errors for estimates. Because PRAMS data also contains information from birth certificate data, there is basic information on women who did not respond to the survey, which allowed the research team to further refine the weights. A discussion of the methods can be accessed at http://www.cdc.gov/prams/methodology.htm.
Strengths:	There are multiple years of data available. Sample size appears to be sufficient for AI/AN analyses.

Pregnancy Risk Assessment Monitoring System (PRAMS)
(continued)

Limitations:	Only 37 states plus New York City and 1 tribe participate in the PRAMS data collection effort, thus impacting the generalizability of estimates to the national level.
Access Requirements and Use Restrictions:	Data are available to the public through a data use agreement at no cost. A research proposal must be mailed or sent electronically to: Denise D'Angelo, MPH Applied Sciences Branch MS-K22, Division of Reproductive Health, Centers for Disease Control and Prevention 4770 Buford Hwy, NE Atlanta GA 30341-3724 DDAngelo@cdc.gov Proposal guidelines and review processes are available at: http://www.cdc.gov/prams/.
Contact Information:	PRAMS website: http://www.cdc.gov/prams. CDC/Division of Reproductive Health 4770 Buford Hwy, NE MS K-20 Atlanta, GA 30341-3717 (770) 488-5200

Resource and Patient Management System (RPMS) and National Patient Information Reporting System (NPIRS)

Sponsor: U.S. Department of Health and Human Services (DHHS)/Indian Health Service (IHS)

Description: The Resource and Patient Management System (RPMS) is an IHS-wide system designed to provide detailed and comprehensive clinical and administrative information to providers and managers at all levels of the Indian health system in order to allow them to better manage individual patients, local facilities, regional and national programs. It has several components for reporting detailed information on patient characteristics, diagnoses, and specific services provided to those patients. RPMS is a decentralized automated information system of over 50 integrated software applications with separate, individual databases at local sites. RPMS software modules fall into three major categories: (1) administrative applications that perform patient registration, scheduling, billing, and linkage functions; (2) clinical applications that support various healthcare programs within IHS; and (3) infrastructure applications. It has the capability to produce special reports, by individual provider, clinic, outpatient versus inpatient services, in addition to other output generated from patient-level records. Taken together, the RPMS components collect, store, and then display an extensive abstract of clinical and administrative information gathered during patient contacts.

A smaller subset of this abstracted information is exported to the National Patient Information Reporting System (NPIRS), a national data warehouse designed to allow IHS to aggregate RPMS data from all their local sites to track clinical practice patterns and episodes of care, provide measures of quality of care and clinical outcomes, perform epidemiological studies, report on patient demographics and healthcare utilization patterns and provide data from which health care costs can be estimated. Data elements exported to NPIRS include certain patient demographics; encounter-based information such as the date, location of a visit (facility), provider, the "Purpose(s) of Encounter" using International Classification of Disease (ICD-9) codes, medications, and certain laboratory test data; and specific patient related clinical data such as health factors.

Relevant Policy Issues: Measurement of Health Status, Disease-specific Measurements, and Factors Contributing to Measured Health Disparities.

Data Type(s): Program enrollment data

Unit of Analysis: Individual

Identification of AI/AN/NA: In the reports that are requested from NPIRS, one can select only AI/AN individuals. The data also permit analyses of a variety of subpopulations, selected by geographic and other variables such as state, reservation, community, facility, tribal affiliation, gender, age group, etc. Tribes have the right to disapprove the release of data that would allow the identification of their tribe.

Resource and Patient Management System (RPMS) and National Patient Information Reporting System (NPIRS) (continued)

AI/AN/NA Population in Data Set:	RPMS system is in use at essentially all Indian Health Service facilities and at many tribal and some urban program sites, and therefore these local databases should have more than sufficient observations to facilitate detailed analyses. Most data pertains to AI/AN patients, although some is about non-AI/AN who obtain care at IHS, tribal, or urban sites for various reasons. NPIRS contains data on all RPMS registered patients and their encounters, but only for a specified subset of their RPMS data. NPIRS also contains similar data from a handful of tribal and/or urban sites who export data to NPIRS in an appropriate format.
AI/AN/NA Subpopulations:	Tribes have the right to approve data release with detailed subpopulation identifiers. Given tribal approval, one could examine the following subpopulations: American Indian alone, Alaska Native alone, or specific tribes and villages.
Geographic Scope:	IHS and tribally operated health care facilities are located in 35 states, while a handful of other states host urban Indian health programs. Geographic areas are identified by state and community. Geographic analysis is available by state, or each individual's tribal affiliation (pending approval by tribe during data request).
Date or Frequency:	Data are continually fed into the RPMS system as patients are served. At sites where data entry into RPMS is performed by clerks from paper encounter forms that providers complete, there can be delays in this data entry that range from days to months. Data from the local RPMS systems are periodically exported to NPIRS. The frequency of these exports from local sites can vary from daily at the largest sites to once a year from a few smaller sites.
Data Collection Methodology:	Participating IHS facilities and providers implement the RPMS software system and input data into the system. Data are then linked into the broader IHS database which can be tapped for research purposes.
Participation:	Optional, with incentives for programs (technical support provided)
Strengths:	Data sets contain a large number of AI/AN/NA respondents. Data are collected on key policy issues including health. There are multiple years of data available. The RPMS data source is a very powerful tool for examining detailed health and utilization information for individuals using the IHS system over time. It contains comprehensive encounter data not otherwise collected through surveys and includes most IHS providers.
Limitations:	RPMS can only report on patients who use IHS facilities and providers and therefore may have some gaps in the overall health experience and utilization of AI/AN patients.
Access Requirements and Use Restrictions:	The data are not available to those outside the agency in raw form, but users can request special data analyses. NPIRS is almost entirely outsourced (provided by a contractor under a contract with IHS), so depending on the scope and complexity of the request the user may or may not have to pay any associated costs. Users

can send data request forms to the Statistics Program at the Office of Program Support at the Indian Health Service (see address below).

Contact Information:

Statistics Program
Office of Program Support
Office of Public Health
Indian Health Service
12300 Twinbrook Parkway, Suite 450
Rockville, Maryland 20852
Telephone: (301) 443-1180
Fax: (301) 443-4087

The RPMS help desk for technical support with the RPMS system has staff who may be able to direct researchers to documents on the website and provide general information about RPMS data. However, the primary purpose of this help desk is for those who are implementing the RPMS system in their facility. The help desk can be contacted by telephone at (505) 248-4371 or (888) 830-7280, or by email to support@ihs.gov.

Main Website, geared primarily towards those who are implementing the RPMS system:
http://www.ihs.gov/Cio/RPMS/index.cfm?module=home&option=index.

Runaway and Homeless Youth Management Information System (RHYMIS)

Sponsor: U.S. Department of Health and Human Services (DHHS)/Administration for Children and Families (ACF)

Description: The Runaway and Homeless Youth Management Information System (RHYMIS) was designed to provide comprehensive information on youth served, issues that affect them and services provided for Runaway and Homeless Youth programs funded by the Family and Youth Services Bureau (FYSB). FYSB mandates that certain data be regularly collected and reported by its grantees. Current grantees must report on the profile of the youth and families they serve, and provide an overview of the services which they deliver under their grant programs. In order to assist grantees in their reporting responsibilities, FYSB funded the development of a Runaway and Homeless Youth Management Information System (RHYMIS).

Relevant Policy Issues: Measures of Well-being for Children, Factors Contributing to Well-being Disparities of Children, Identification of Evidence-based Practices and Programs that Improve Child Well-being and are Generalizable/Replicable, and Homelessness.

Data Type(s): Program reporting data

Unit of Analysis: Individual

Identification of AI/AN/NA: Race and ethnicity are self reported by the youth. This sometimes results in multiple indications or "not provided" as responses. Below are the instructions provided to program staff on the data collection forms:

"How does the youth describe himself/herself using these census categories? On the basis of the youth's self-perception, select one or more codes indicating the young person's race category and one code indicating their ethnicity category."

<u>The race categories are:</u>
American Indian or Alaska Native (AI/AN): A person having origins in any of the original peoples of North and South America (including Central America) and who maintains tribal affiliation or community attachment.

Asian: A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.

Black or African American: A person having origins in any of the black racial groups of Africa.

Native Hawaiian or other Pacific Islander (NH/PI): A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.

White: A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.

AI/AN/NA Population in Data Set:

For FY 2004:
TOTAL number of records in data set: 56,677
AI/AN: 1,922
NH/PI: 338
AI/AN and some other race(s): 398
NH/PI and some other race(s): 47

For FY 2003:
TOTAL number of records in data set: 74,290
AI/AN: 2,497
NH/PI: 581
AI/AN and some other race(s): 481
NH/PI and some other race(s): 78

Geographic Scope: The geographic scope of RHYMIS is national. State and regional identifiers are included in the data; therefore geographic analysis is possible at the state and regional levels. Although federally funded programs within all states submit data to RHYMIS, the data should not be used to produce state-wide estimates of runaway and homeless youth. RHYMIS data are only collected on youth who utilize federally funded programs under the Runaway and Homeless Youth Act. However, the data can be used to create state-level estimates of youth who use federally funded programs.

Date or Frequency: RHYMIS is collected on a semi-annual basis. Due dates are listed as April 15th and October 15th. Data are released for a full fiscal year. Prior to FY 2002, sometimes fewer than 45-55 percent of participating programs reported fully to RHYMIS. For this reason, use of data prior to FY 2002 is discouraged by RHYMIS staff. RHYMIS was re-designed during FY 2001; as a result, all data beginning in FY 2002 is based on virtually a 100 percent response rate.

Data Collection Methodology: Program staff utilize desktop software provided by FYSB to complete an intake and exit form for each youth serviced by the program.

Participation: Mandatory

Authorization: Authorization for RHYMIS falls under the Runaway and Homeless Youth Act as Reauthorized (2003) by 42 U.S.C. 5701.

Strengths: This data set contains a large number of AI/AN/NA respondents. Data are collected on key policy issues, including (after 2004) history of involvement in the child welfare system (e.g., foster care experiences). Multiple years of complete data are available. Starting with the date range October 1, 2001 to the end of the most recent fiscal year (September 30, 2005), the data are virtually a 100 percent complete report on all youth served by all FYSB-funded runaway and

homeless youth agencies. Generally less than 1 percent of the grantees fail to report at all by the time the database is closed for a six-month reporting period.

Limitations: There is little documentation available on the RHYMIS data. Studies that have used the RHYMIS data may discuss issues related to data quality, but ACF does not release a methodology report for data users.

RHYMIS data should not be used to generate estimates for all runaway and homeless youth. RHYMIS data focus on youth who are served by federally funded programs under the Runaway and Homeless Youth Act.

Access Requirements and Use Restrictions: NEO-RHYMIS is an online reporting system that contains complete data at the FYSB grantee level. Researchers interested in detailed RHYMIS research should contact the RHYMIS Hotline at (888) 749-6474. There are some limitations on use involving privacy issues.

Contact Information:

Director, Division of Research and Evaluation
Family and Youth Services Bureau
www.acf.hhs.gov/programs/fysb
phone: (202) 205-8496; fax: (202) 690-5600

Arlene Calabro, RHYMIS Support
acalabro@csc.com
(954) 472-4122
Computer Sciences Corporation
15245 Shady Grove Road
Rockville, MD 20850

Small Area Income and Poverty Estimates (SAIPE)

Sponsor: U.S. Department of Commerce/U.S. Census Bureau

Description: The U.S. Census Bureau, with support from other federal agencies, created the Small Area Income and Poverty Estimates (SAIPE) program to provide more current estimates of selected income and poverty statistics than the most recent decennial census. Estimates are created for states, counties, and school districts. The main objective of this program is to provide updated estimates of income and poverty statistics for the administration of federal programs and the allocation of federal funds to local jurisdictions.

The estimates are not direct counts from enumerations or administrative records, nor direct estimates from sample surveys. Data from those sources are not adequate to provide intercensal estimates for all counties. Instead, the relationship between income or poverty and tax and program data for the states and a subset of counties are modeled using estimates of income or poverty from the Annual Social and Economic Supplement (ASEC) to the Current Population Survey (CPS). The modeled relationships are then used to develop estimates for all states and counties. For school districts, the model-based county estimates and the decennial census distribution of the population in poverty of each county across its constituent school districts are used to create the estimates.

Relevant Policy Issues: Income status.

Data Type(s): Statistical database

Unit of Analysis: Estimates developed for the SAIPE program are not at the individual level, therefore counts of the AI/AN/NA population in the data set are not available. The estimates, however, are available by county and school district (geographical units that may be of interest). Researchers could use other data sources to identify geographic areas with large concentrations of AI/AN/NA and then use SAIPE data to do analyses of these areas.

Identification of AI/AN/NA: AI/AN/NA individuals are not identified in the data set.

Geographic Scope: The geographic scope of the study is national. Geographic analysis is possible by state, county, and school district.

Date or Frequency: State and county data are available for 1989, 1993, and 1995 – 2003.
School district data are available for 1995, 1997, and 1999 – 2003.

Data Collection Methodology: No data are collected. Models are developed on a periodic basis and then used to generate estimates.

Authorization: The SAIPE program was developed when Congress called for authorization legislation requiring the Secretary of Commerce to develop the methodology to

produce "intercensal data relating to the incidence of poverty for each state, county, and local jurisdiction." The legislation further called for estimates of the number of children impovered age 5 to 17, for local education agencies (school districts) and of the number of impovered people age 65 and over for states and counties. In September 1994, the Congress passed the Improving America's Schools Act and signed it into law (PL 103-382). It reauthorized and amended the Elementary and Secondary Education Act. Authorization for SAIPE falls under this legislation. The No Child Left Behind Act of 2001 further amended the ESEA and required annual production of estimates for school districts.

Strengths: The SAIPE program data provides estimates of income and poverty statistics based on more current data than other sources of information.

Limitations: The type of information available from SAIPE is not diverse; it concerns only the number in poverty, poverty rates, and median household income information. SAIPE estimates are based on statistical models and are subject to modeling error.

Other: The Small Area Health Insurance Estimates (SAHIE) program builds on the work of the SAIPE program. SAHIE was created to develop model-based estimates of health insurance coverage by age for counties and states. The SAHIE program has developed experimental estimates for counties and states for 2000 for the total population with and without health insurance coverage; children under age 18 with and without health insurance coverage; and measures of uncertainty of the estimates. This type of county-level data on health insurance coverage are not available elsewhere because neither the decennial census nor the American Community Survey contain questions on this topic. More information about the SAHIE program can be found at:
http://www.census.gov/hhes/www/sahie/index.html

Access Requirements and Use Restrictions: These estimates are available to the public at no cost.

Contact Information: Mail address:
U.S. Census Bureau
4700 Silver Hill Road
Washington DC 20233-0001

Telephone:
For general questions about SAIPE, contact the Statistical Information Staff of the Data Integration Division at this phone number: (301) 763-3242.

Location of the actual data:
The data are available at the following website:
http://www.census.gov/hhes/www/saipe/tables.html
These data are also available via DataFerrett

Reports of Interest:

Detailed information on SAIPE estimates and limitations concerning the use of these data are provided in three recent publications:

1. *Evaluation of School District Poverty Estimates: Predictive Models using IRS Income Tax Data.* Jerry J. Maples and William R. Bell, Bureau of the Census, Washington DC 20233.
(http://www.census.gov/hhes/www/saipe/asapaper/asa05finalmaples.pdf)

2. *Using Medicaid Participant Data in the Estimation of County Poverty Levels.* David S. Powers, U.S. Census Bureau, Small Area Estimates Branch, Housing and Household Economic Statistics Division, Room 1451, Building 3.
(http://www.census.gov/hhes/www/saipe/asapaper/asa2005dpowers.pdf)

3. *Estimating School District Poverty with Free and Reduced Lunch Data.* Craig Cruse and David Powers, U.S. Census Bureau, Small Area Estimates Branch, Room 1451-3.
http://www.census.gov/hhes/www/saipe/publications.html.

Surveillance, Epidemiology, and End Results (SEER)

Sponsor: U.S. Department of Health and Human Services (DHHS)/National Institutes of Health (NIH)/National Cancer Institute (NCI)

Description: The Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute is responsible for the collection and reporting of cancer incidence and survival data from 15 population-based central cancer registries that cover 26 percent of the U.S. population. The U.S. racial/ethnic population coverage in SEER includes 23 percent of African Americans, 40 percent of Hispanics, 42 percent of American Indians and Alaska Natives, 53 percent of Asians, and 70 percent of Native Hawaiian and other Pacific Islanders. SEER data include patient demographic information as well as primary tumor site, tumor morphology and stage at diagnosis, first course of cancer treatment, and follow-up for vital status. SEER began collecting data on cancers diagnosed on January 1, 1973, which enables the analysis of longitudinal trends as well as current patterns of cancer.

Relevant Policy Issues: Disease-specific Measurements and Health Disparities.

Data Type(s): Registry

Unit of Analysis: Cancer case (may be more that one cancer diagnosis per person in the database).

Identification of AI/AN/NA: Detailed racial/ethnic information is collected for over 30 different racial/ethnic categories including, but not limited to:

- White
- Black
- American Indian/Alaska Native (AI/AN)
- Asian or Pacific Islander (e.g., Chinese, Japanese, Filipino, Native Hawaiian, Korean, …others).
- Hispanic/Latino

AI/AN/NA Population in Data Set: The SEER database includes information on over six million in situ and invasive cancer cases with more than 350,000 cases being added each year. Of these, over 28,000 cases are among AI/ANs. Geographic regions with large AI/AN populations in the publicly available SEER data include New Mexico, Alaska, California and the Seattle/Puget Sound area.

AI/AN/NA Subpopulations: Tribal affiliation is not reported in SEER data, but geographic-specific data analyses may better characterize cancer patterns in specific AI/AN subpopulations.

Geographic Scope: SEER covers geographically and demographically diverse populations in the U.S. including all residents of the states of CA, CT, HI, IA, KY, LA, NJ, NM, and UT; metropolitan areas of Atlanta, Detroit, Seattle; selected rural Georgia counties;

and AI/AN populations in AK and AZ. Available geographic identifiers within the database include registry (which covers either a state or a group of counties) and county.

Date or Frequency: Data are available on an annual basis from 1973 to the present.

Data Collection Methodology: Population-based cancer registries from state or metropolitan area or rural county grouping submit data to the National Cancer Institute for inclusion in the SEER database. The cancer patient data are collected from health providers such as hospitals, clinics, pathology labs, and physician offices as well as from autopsy reports and death certificates. The data are subjected to rigorous data quality edits and investigations and must meet data quality standards. The SEER Program data are considered the international standard for cancer registry data quality.

Participation: Optional, without incentives. The population-based registries report their data through contracts or interagency agreements with the NCI.

Strengths: Registries contain a large number of AI/AN/NA respondents including a region that is predominantly Alaska Native. Data are collected on key policy issues including health; for example, detailed data are collected on cancer type, stage, morphology, first course of treatment, survival, cause of death and patient demographics. There are multiple years of data available.

Limitations: The SEER data are a definitive source of cancer incidence and survival data in the U.S., but coverage is limited to about 26 percent of the total U.S. population. Minority racial/ethnic groups, foreign-born, and urban populations are groups of special interest to the SEER program and are therefore somewhat overrepresented in the database. Although frequency distributions of tumor characteristics and observed survival may be generated for over 30 detailed racial/ethnic groups, incidence rate calculations are limited to the racial/ethnic groups for which population denominators are available from the Census Bureau. Incidence rates for AI/AN can be calculated for diagnoses in 1992 and later.

Although AI/AN are well-represented in the database, the three states (AZ, CA, and AK) for which data for AI/AN are primarily collected are not necessarily representative of all AI/AN, since there is evidence that cancer incidence may be different in geographically distinct AI/AN populations.

Other: Since 1994, the National Program of Cancer Registries (NPCR) has been funding state cancer registries to collect population-based cancer incidence data. Starting in 2001, NPCR began receiving data annually from funded programs with the goals of establishing the quality of the data and eventually releasing the data for use in public health planning. Currently, the United States Cancer Statistics on the NPCR website (www.cdc.gov/cancer/npcr/uscs) provides aggregate rates for states by race. Information on AI/AN cancer incidence is available for some states that meet the 100,000 population criteria through this analysis system.

NPCR plans to have a restricted use dataset available for use by researchers who meet specified criteria. Some NPCR registries provide county-level data on AI/AN to State Cancer Profiles (http://statecancerprofiles.cancer.gov).

Access Requirements and Use Restrictions:

Potential users must sign a data use agreement (http://seer.cancer.gov/publicdata/access.html). Data tables are available without any data use agreement requirements.

Contact Information:

Cancer Statistics Branch
Surveillance Research Program
Division of Cancer Control and Population Sciences
National Cancer Institute
Suite 504, MSC 8316
6116 Executive Boulevard
Bethesda, MD 20892-8316
(301) 496-8510

Information on SEER public use data is available at the following website: http://seer.cancer.gov/publicdata/

Questions can be addressed to seerweb@imsweb.com.

Survey of Jails in Indian Country (SJIC)

Sponsor:	U.S. Department of Justice/Bureau of Justice Statistics
Description:	The Survey of Jails in Indian Country (SJIC), a component of the Annual Survey of Jails, gathers data on all adult and juvenile jail facilities and detention centers in Indian Country, which is defined as reservations, pueblos, rancherias, and other Native American and Alaska Native communities throughout the United States. The survey, conducted yearly between 1998 and 2004, is a complete enumeration of all confinement facilities operated by tribal authorities or the Bureau of Indian Affairs (BIA) and provides data on number of inmates and facility characteristics and needs. Variables describe each facility, including capacity, number of adult inmates, number of juveniles held, number of inmates held by sex and conviction status, number of admissions and discharges in the last 30 days, number of inmate deaths, the peak population during June, facility crowding, and renovation and building plans. The 2004 survey also collected information on inmate health services and programs available to inmates including information on four infectious diseases, including HIV, hepatitis B and C, and tuberculosis. Additional new information included inmate medical and mental health services, suicide prevention, substance dependency programs, domestic violence counseling, sex offender treatment, educational programs, and inmate work assignments.
Relevant Policy Issues:	Justice System Issues.
Data Type(s):	Survey
Unit of Analysis:	Correctional facility
Identification of AI/AN/NA:	American Indian and Alaska Native (AI/AN) individuals are not identified in this study. Instead, this study is a complete enumeration of all jails and correctional facilities in Indian Country.
AI/AN/NA Population in Data Set:	In 2001, this study included 68 facilities. In 2002 and 2003, this study included 70 facilities.
Geographic Scope:	The geographic scope of the study includes AI/AN communities. All identifiers for the 70 respondent facilities are included in the data file (i.e., facility name, tribal affiliation, city, state, zip code). While facilities from 19 different states and 55 different tribes participate, geographic analysis would not be appropriate given the small number of facilities in any one tribe or state.
Date or Frequency:	Data were collected annually from 1998-2004.

Survey of Jails in Indian Country (SJIC)
(continued)

Data Collection Methodology: The survey was conducted by mail. Surveys were mailed to each facility and facility-identified staff completed the surveys. Data were returned by mail, fax, or telephone.

Participation: Optional, without incentives.

Response Rate: Through follow-up phone calls and facsimiles, the 2002 survey achieved an 86 percent response rate. Older data for non-responding facilities is included in reports released by Bureau of Justice Statistics.

Strengths: There are multiple years of data available providing institutional-level descriptions of the conditions of confinement in Indian Country.

Limitations: These are not individual-level data; they are a description of facilities. Moreover, in this survey, race of persons being confined is not asked so it is not possible to determine how many persons described in the data are AI/AN.

Access Requirements and Use Restrictions: Data are available to public at no cost.

Contact Information: Data archive information:
National Archive of Criminal Justice Data
ICPSR
University of Michigan
Institute for Social Research
P.O. Box 1248
Ann Arbor, MI 48106-1248
(800) 999-0960
(313) 763-5011
nacjd@icpsr.umich.edu

Questions for the Bureau of Justice Statistics should be mailed to:
Todd Minton
Statistician
Corrections Statistics Program
Bureau of Justice Statistics
810 Seventh St, NW
Washington, DC 20531
(202) 305-9630

Data for 1998-2001 can be downloaded at http://webapp.icpsr.umich.edu/cocoon/NACJD-SERIES/00158.xml. Data for 2002, 2003, and 2004 were not yet available as this catalog was being prepared.

Survey of Program Dynamics (SPD)

Sponsor: U.S. Department of Commerce/U.S. Census Bureau

Description: The Survey of Program Dynamics (SPD) is a longitudinal database drawn from a study designed to collect data on the economic, household, and social characteristics of a nationally representative sample of the U.S. population over time. Core data include employment, income, welfare program participation, health insurance and utilization, child well-being, marital relationships, and parents' depression. The SPD also had topical modules that vary by year. The primary goals of the SPD were to provide information on spells of actual and potential welfare program participation (over a ten-year period), to examine the causes of program participation and its long-term consequences (on recipients and their families), and to monitor the possible long-term changes (for individuals) that result from implementing welfare reform.

Relevant Policy Issues: Measurement of Health Status, Income Status, Unemployment Rates, Economic Assistance Program Participation Rates, Economic Opportunity, Educational Attainment, Factors Contributing to Educational Disparities, Measures of Well-being for Families/households, Factors Contributing to Well-being Disparities of Families, Measures of Well-being for Children, Factors Contributing to Well-being Disparities of Children, Transportation Availability, Eligibility and Use of Other Veterans Administration Benefits.

Data Type(s): Survey

Unit of Analysis: Individual

Identification of AI/AN/NA: The race item on all versions of the SPD questionnaire reads: “Which of these categories best describes (your/name's) race?”

- White
- Black
- American Indian (AI), Aleut or Eskimo
- Asian or Pacific Islander

AI/AN/NA Population in Data Set: In the 1992 - 2002 longitudinal file with a total of 129,013 records, there are 1,100 unique American Indian, Aleut or Eskimo respondents.

Geographic Scope: The geographic scope of the study is national. Geographic areas identified are regions (i.e., Northeast, Midwest, South, West), and states. Geographic analysis is possible by region and by state.

Date or Frequency: The SPD is a longitudinal database. Data were collected annually from 1997 - 2002. There are no plans for future administrations of the SPD.

Survey of Program Dynamics (SPD)
(continued)

Data Collection Methodology: Data are collected in person using computer-assisted personal interviewing (CAPI). The original pool of respondents for the SPD were households that were previously interviewed in the 1992 and 1993 Survey of Income and Program Participation (SIPP) panels. The first round of SPD data collection was conducted in 1997 using a modified version of the March Current Population Survey. After the 1997 wave of data collection, the SPD questionnaire was developed and administered from 1998 through 2002.

Participation: Optional, with incentives

Response Rate: Unweighted response rates are reported for the 1997-2002 administrations of the SPD as follows:

- 1997 SPD: 81.7%
- 1998 SPD: 85.0%
- 1999 SPD: 85.2%
- 2000 SPD: 79.7% (Including non-interviewed households from the 1997 list)
- 2001 SPD: 74.1% (Including non-interviewed households from the 1997 list and the 1992-1993 SIPP)
- 2002 SPD: 65.1%

Sampling Methodology: The 1997 SPD recontacted the sample members who were interviewed for the 1992 and 1993 SIPP panels. The SIPP samples were multistage, stratified samples of the U.S. civilian noninstitutionalized population.

The sample size for the 1997 SPD was 34,609 households. Census field representatives interviewed 30,125 households. At any given point in time, a household was eligible to be interviewed if it contained an original sample member (age 15 or older). The number of eligible households fluctuated from round to round of interviewing because of household formation and dissolution—and because original sample members move from one (previously eligible) household to another (previously ineligible) household.

Strengths: Data are collected on key policy issues, including health and child welfare. There are multiple years of data available. The documentation of the content and implementation of the SPD is comprehensive and available on-line.

Limitations: Asians and Pacific Islanders are collapsed into a single response category in all versions of the SPD.

Access Requirements and Use Restrictions: Data are available to the public at no cost. Data are publicly available for each wave of data collection as well as a longitudinal file across all years of data collection.

Contact Information:

The actual data are available for download from the following website: http://www.bls.census.gov/spd/access.html.

Questions about the SPD should be addressed to: dsd.survey.program.dynamics@census.gov

Temporary Assistance for Needy Families (TANF) and Tribal TANF

Sponsor: U.S. Department of Health and Human Services (DHHS)/Administration for Children and Families (ACF)

Description: The Temporary Assistance for Needy Families (TANF) and Tribal TANF database contains demographic characteristics for families receiving assistance under the TANF program. TANF case record information is reported to the national TANF database by states and territories on a quarterly basis. The database consists of active cases (families who were receiving assistance for the reporting month by the end of the sample month) and closed cases (families whose assistance was terminated for the reporting month, but received assistance in the prior month). States have the option of submitting all active and closed cases or a sample of these cases.

Since 1996, federally recognized American Indian Tribes and Alaska Native organizations have been allowed to operate their own TANF programs and serve tribal members who would otherwise be served by the state in which they live. As of Fiscal Year (FY) 2005 year's end, 51 Tribal TANF plans were approved to operate on behalf of 237 tribes and Alaska Native villages. American Indian and Alaska Native families not served by Tribal TANF programs continue to be served by state TANF programs. The Tribal TANF database includes demographic characteristics of families receiving assistance under Tribal TANF.

Relevant Policy Issues: Income Status, Unemployment Rates, Economic Assistance Program Participation Rates, and Measures of Well-being for Families/Households.

Data Type(s): Program reporting data

Unit of Analysis: Analysis can be conducted at the individual level and family level.

For reporting purposes, the TANF family means a) all individuals receiving assistance as part of a family under the state's TANF Program; and b) the following additional persons living in the household, if not included under a) above: 1) parent(s) or caretaker relative(s) of any minor child receiving assistance; 2) minor siblings of any child receiving assistance; and 3) any person whose income or resources would be counted in determining the family's eligibility for or amount of assistance.

For Tribal TANF, tribes administering their own TANF program have great flexibility in program design and implementation. They can define such elements of their programs as the service area; service population, including the definition of "family"; time limits; benefits and services; and work activities.

Identification of AI/AN/NA: The State TANF agencies or Tribal TANF grantees collect and report data for each person receiving TANF assistance. The instructions for reporting race/ethnicity on TANF recipients are as follows:

The intent of this data element is to capture the multiplicity of race and ethnicity characteristics applicable to each person. States/tribes should code at least one of the race categories "YES" in addition to coding ethnicity.

The provided race/ethnicity categories include:
Ethnicity:

- Hispanic or Latino

Race:

- American Indian or Alaska Native (AI/AN)
- Asian
- Black or African American
- Native Hawaiian or other Pacific Islander (NH/PI)
- White

AI/AN/NA Population in Data Set:

Researchers can receive a sample of the FY 2004 state TANF database for research. This database contains 205,119 records for active cases, and 58,453 records for closed cases. Breakdowns of the AI/AN/NA population are below:

Active Cases
AI/AN: 9,718
NH/PI: 1,711

Closed Cases
AI/AN: 3,001
NH/PI: 653

In 2002 there were 9,983 families receiving Tribal TANF assistance in total. The numbers for the 2004 database have not been released yet.

AI/AN/NA Subpopulations:

Researchers can request analyses of Tribal TANF data by tribal affiliation. Because the numbers in the data set are very small, extreme limitations will be put on such requests.

Geographic Scope:

The geographic scope of the state and Tribal TANF databases is national. In the state TANF database, analysis by state is possible. Also, in the state TANF database, the 3-digit county Federal Information Processing Standards (FIPS) code is provided. County-level analysis may be possible. Tribal TANF grantees do not report the FIPS code, but they do report the 3-digit tribal identification code instead. Tribal-level special tabulations may be available, but due to the small numbers in the data, extreme limitations will be put on requests.

Date or Frequency:

Data is collected on a monthly basis and submitted quarterly to the national TANF databases. Research databases are compiled for an entire fiscal year. The most current fiscal year data available is FY 2004.

Temporary Assistance for Needy Families (TANF) and Tribal TANF (continued)

Data Collection Methodology:	State and Tribal TANF agencies complete a TANF data collection form for all families receiving assistance under the TANF program.
Participation:	Mandatory
Sampling Methodology:	There is no single sampling method applied across the board for all states submitting data to the national TANF database. Twenty-nine states submitted records on all active and closed cases, while the remaining 24 states submitted sample data. If states do not meet the annual minimal sample size requirements, they must report data for all active and closed cases. No tribe has a caseload large enough to warrant sampling.
Authorization:	The Personal Responsibility and Work Opportunity Reconciliation Act of 1996 requires states, territories, and tribes to collect on a monthly basis and report to the Secretary of the Department of Health and Human Services on a quarterly basis disaggregated case record information on families receiving assistance, families no longer receiving assistance, and families newly-approved for assistance from programs funded under TANF.
Strengths:	Data sets contain a large number of AI/AN/NA respondents though Tribal TANF data are not currently available. Multiple years of TANF data are available.
Limitations:	There is limited documentation available for researchers who wish to use the TANF public use database.
Access Requirements and Use Restrictions:	Tribal TANF is still relatively new. Processes to provide researchers with direct access to the data files may be developed in the future. For the time being, the office is concerned about the confidentiality of the data, but is willing to run analyses for researchers. Also, it may be possible to request the data with tribal affiliation removed, although detail would be lost.
Contact Information:	Researchers interested in receiving the sample FY 2004 state TANF public use database should contact: Andrew Yoo (202) 401-5098 AYoo@acf.hhs.gov Researchers interested in working with Tribal TANF data should contact the Tribal TANF acting director Bob Shelbourne at (202) 401-5150; Raymond Apodaca, Tribal TANF Team Leader at (202) 401-5150; Ann Bowker, Native Employment Works (NEW) Program at (202) 401-5308; or Gerald Joireman, TANF Data at (202) 401-5097, email: gjoireman@acf.hhs.gov.
Reports of Interest:	The 2004 TANF/TTANF Annual Report to Congress can be located at: http://www.acf.hhs.gov/programs/ofa/indexar.htm

Tobacco Use Supplement to the Current Population Survey (TUS-CPS)

Sponsor: U.S. Department of Health and Human Services (DHHS)/National Institutes of Health/National Cancer Institute (NCI) and The Centers for Disease Control and Prevention (CDC)

Description: The Tobacco Use Supplement to the Current Population Survey (TUS-CPS) is a survey of tobacco use that serves as a key source of national and state level data on smoking and other tobacco use in the U.S. household population. The TUS-CPS uses a large, nationally representative sample that contains information on about 240,000 individuals within a given survey period. Although the TUS-CPS has changed slightly between 1992 and 2003, it has generally contained about 40 items concerning cigarette smoking prevalence including smoking history, current and past cigarette consumption; cigarette smoking quit attempts and intentions to quit; medical and dental advice to quit smoking; cigar, pipe, chewing tobacco, and snuff use; workplace smoking policies; smoking rules in the home; attitudes toward smoking in public places; opinions about the degree of youth access to tobacco in the community; and attitudes toward advertising and promotion of tobacco. These data can be used by researchers to monitor progress in the control of tobacco use, conduct tobacco-related research, and evaluate tobacco control programs.

Relevant Policy Issues: Key Health Disparities of Priority Interest.

Data Type(s): Survey

Unit of Analysis: Individual

Identification of AI/AN/NA: The race/ethnicity of the TUS-CPS respondents is taken from the CPS data. In the CPS, participants are asked to respond to the question on race by indicating one or more of six race categories. The six race categories are:

- White
- Black or African American
- American Indian/Alaska Native (AI/AN)
- Asian
- Native Hawaiian/Other Pacific Islander (NH/PI)
- Some Other Race (this category is not read or displayed to the respondent)

Responses to the race item are recoded into multiple race categories for analytic purposes:

- AI/AN Only
- NH/PI Only
- White/AI/AN
- White/NH/PI
- Black/AI/AN Black/NH/PI

- AI/AN/Asian
- Asian/NH/PI
- White/Black/AI/AN
- White/AI/AN/Asian
- White/Asian/NH/PI
- White/Black/AI/AN/Asian

AI/AN/NA Population in Data Set:

For the TUS-CPS, some surveys are completed with proxy respondents when the original sampled respondent is unavailable. Self-respondents are eligible for the entire TUS-CPS questionnaire, whereas proxy respondents are only eligible for certain items. Additionally, responses to the race item are recoded into the multiple race categories. The following categories reflect the unweighted counts for AI/AN/NA respondents, including all surveys completed by the sampled respondents and proxies, and all surveys completed by only the sampled respondent, in the February, June, and November 2003 CPS:

February 2003 (N = 68,954)
AI/AN Only: 786 self & proxy, 582 self only
NH/PI Only: 218 self & proxy, 152 self only
White/AI/AN: 660 self & proxy, 552 self only
White/NH: 52 self & proxy, 34 self only
Black/AI/AN: 71 self & proxy, 59 self only
Black/NH/PI: 3 self & proxy, 2 self only
AI/AN/Asian: 8 self & proxy, 3 self only
Asian/NH/PI: 36 self & proxy, 23 self only
White/Black/AI/AN: 29 self & proxy, 21 self only
White/AI/AN/Asian: 1 self & proxy, 1 self only
White/Asian/NH/PI: 1 self & proxy, 1 self only
White/Black/AI/AN/Asian: 2 self & proxy, 2 self only

June 2003 (N = 89,864)
AI/AN Only: 963 self & proxy, 739 self only
NH/PI Only: 254 self & proxy,164 self only
White/AI/AN: 784 self & proxy, 629 self only
White/NH: 72 self & proxy, 46 self only
Black/AI/AN: 59 self & proxy, 46 self only
Black/NH/PI: 3 self & proxy, 1 self only
AI/AN/Asian: 3 self & proxy, 3 self only
Asian/NH/PI: 72 self & proxy, 36 self only
White/Black/AI/AN: 40 self & proxy, 34 self only
White/AI/AN/Asian: 5 self & proxy, 4 self only
White/Asian/NH/PI: 6 self & proxy, 3 self only
White/Black/AI/AN/Asian: 0 respondents

November 2003 (N = 90,802)
AI/AN Only: 931 self & proxy, 672 self only
NH/PI Only: 222 self & proxy, 142 self only
White/AI/AN: 827 self & proxy, 629 self only
White/NH: 58 self & proxy, 28 self only
Black/AI/AN: 72 self & proxy, 55 self only
Black/NH/PI: 3 self & proxy, 2 self only
AI/AN/Asian: 5 self & proxy, 2 self only
Asian/NH/PI: 90 self & proxy, 54 self only
White/Black/AI/AN: 36 self & proxy, 24 self only
White/AI/AN/Asian: 3 self & proxy, 1 self only
White/Asian/NH/PI: 7 self & proxy, 5 self only
White/Black/AI/AN/Asian: 0 respondents

Geographic Scope: The geographic scope of the study is national. Due to the large sample size for most survey items, analyses can be done at either the national or state levels, and in some cases, for areas smaller than the state level. State sample sizes range from 2,100 for the District of Columbia to 18,700 for California. State data for any year is considered most reliable when using data from all 3 months of data collection.

Date or Frequency: The TUS-CPS was administered as part of the CPS in 1992-1993, 1995-1996, 1998-1999, 2000, 2001-2002, and 2003. For these time periods, the TUS-CPS was administered for 3 months throughout the year.

Over the next 10 years, NCI plans to conduct the TUS-CPS triennially, alternating between a core questionnaire intended for monitoring purposes (similar to the questionnaire used throughout the 1990s) and more specific Special Topics questionnaires that target tobacco-related issues of particular interest to researchers. NCI and CDC will be co-sponsoring the supplements. The next round of core TUS-CPS supplements is being fielded in May 2006, August 2006, and January 2007.

Aggregation: It is recommended that when analyzing the TUS-CPS data, researchers should aggregate the data across the months of the data collection effort for a single year. For example, when using the 2003 data researchers should combine the data collected in the months of February, June, and November.

Although multiple years of TUS-CPS data are available, in 2003 significant changes were made to the race/ethnicity questions in the CPS. In 2003, respondents were able to select more than one race when answering the survey. This change in wording does not impact smoking estimates and trends calculated for the entire nation from the TUS-CPS, but it could potentially impact smoking estimates and trends calculated by race/ethnicity. NCI has developed a method to construct single-race estimates using data from the post-2003 TUS-CPS. The method is useful when trends over time are being examined for single race groups using both pre-2003 and post-2003 data. More information is available in the report *Bridging Estimates by Race for the Tobacco Use Supplement to the*

Current Population Survey – (TUS-CPS) (http://riskfactor.cancer.gov/studies/tus-cps/race_bridging.pdf) which describes the method and gives an initial assessment of the usefulness of the race adjustment.

Data Collection Methodology: The mode of data collection is both telephone and in-person interviewing. NCI estimates that 75 percent of the respondents reply to the survey by telephone and 25 percent of the respondents reply during personal home visits. Additionally, about 20 percent of the completed surveys are completed by a proxy for the sampled respondent. When a proxy is providing the information, only a few measures of use are collected.

Participation: Optional, without incentives

Response Rate: Nonresponse rates are less than 9 percent for the monthly CPS for September 2003 through September 2004.

Sampling Methodology: The TUS-CPS was conducted as a supplemental study with the core CPS. The CPS sample is a multistage stratified sample of approximately 56,000 housing units from 792 sample areas. The CPS samples housing units from lists of addresses obtained from the 1990 Decennial Census of Population and Housing. These lists are updated continuously for new housing built after the 1990 census. The first stage of sampling involves dividing the United States into primary sampling units (PSUs)—most of which comprise a metropolitan area, a large county, or a group of smaller counties. Every PSU falls within the boundary of a state. The PSUs are then grouped into strata.

Analysis: Effective sample size, design effects, and standard errors for estimates are discussed in detail in the following publication: *Technical Document CPS03: Current Population Survey, February, June, and November 2003: Tobacco Use Supplement File* (http://riskfactor.cancer.gov/studies/tus-cps/surveys/cps03_tech_doc.pdf).

Strengths: There are a large number of AI/AN/NA respondents. There is relevance to a key health policy issue, tobacco use. Multiple years of data are available.

Limitations: While multiple years of data are available, in 2003 significant changes were made to the race/ethnicity questions in the CPS that may affect the ability to look at tobacco use by AI/AN/NA persons over time.

Access Requirements and Use Restrictions: TUS-CPS data are available to the public. The data may be purchased through the Census Bureau's online catalog. Prices for the data may vary; the cost of purchasing the 2003 data on CD-ROM is $55.00.

Tobacco Use Supplement to the Current Population Survey (TUS-CPS)
(continued)

Contact Information:	Risk Factor Monitoring and Methods Branch Applied Research Program Division of Cancer Control and Population Sciences National Cancer Institute, EPN 4005 6130 Executive Blvd-MSC 7344 Bethesda, MD 20892-7344 (301) 496-8500 Instructions for ordering the TUS-CPS data files are available on the NCI website at: http://riskfactor.cancer.gov/studies/tus-cps/info.html
Reports of Interest:	U.S. Department of Health and Human Services. *Tobacco Use Among U.S. Racial/Ethnic Minority Groups—African Americans, American Indians and Alaska Natives, Asian Americans and Pacific Islanders, and Hispanics: A Report of the Surgeon General.* Atlanta, Georgia: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 1998. http://www.cdc.gov/TOBACCO/sgr/sgr_1998/index.htm

Treatment Episode Data Set (TEDS)

Sponsor: U.S. Department of Health and Human Services (DHHS)/Substance Abuse and Mental Health Services Administration (SAMHSA)

Description: The Treatment Episode Data Set (TEDS) is an administrative data system providing descriptive information about the national flow of admissions to providers of substance abuse treatment. The TEDS series was designed to provide annual data on the number and characteristics of persons admitted to public and private substance abuse treatment programs receiving public funding. Data collected include client demographics, client substance abuse problems, client mental health information, information on treatments received and source of client referral to treatment, and sources of payment for treatment. Admission data have been collected since 1989. In 2000, a discharge data set was added to allow TEDS to collect information on entire treatment episodes. TEDS is comprised of data that are routinely collected by states in monitoring substance abuse treatment facilities. In general, TEDS data cover those facilities that receive state funds for substance abuse treatment.

Relevant Policy Issues: Health Disparities and Differences in Patterns in Drug and Alcohol Use.

Data Type(s): Program reporting data

Unit of Analysis: Admissions at publicly funded substance abuse treatment facilities

Identification of AI/AN/NA: Data are reported in the following categories:

- Alaska Native (Aleut, Eskimo, Indian) (AN)
- American Indian (Other than Alaskan Natives) (AI)
- Asian or Pacific Islander
- Black
- White
- Other Single Race
- Two or More Races

AI/AN/NA Population in Data Set: In 2004 the entire TEDS data include 1,875,026 cases. Counts for AI/AN cases are:
AN: 5,186
AI: 38,785

AI/AN/NA Subpopulations: American Indian alone and Alaska Native alone are available.

Geographic Scope: The geographic scope of TEDS is national. Geographic indicators include state, primary metropolitan statistical area (PMSA), metropolitan statistical area (MSA), and core-based statistical area (CBSA).

Treatment Episode Data Set (TEDS)
(continued)

Date or Frequency: TEDS data are compiled yearly. Data for 1992-2004 are available online. New files will continue to be released approximately 18 months after the end of each year (e.g., the 2004 file was released in June 2006.)

Data Collection Methodology: TEDS data are routinely collected by state administrative systems and then submitted to SAMHSA in a standard format.

Participation: Participation is mandatory for publicly-funded clients. Other clients participate on an optional basis.

Response Rate: TEDS is designed to include client-level data from all facilities that receive state funds for substance abuse treatment. In 1997, the most recent information available, TEDS was estimated to represent 83 percent of all admissions to these facilities. Also in 1997, TEDS was estimated to cover 67 percent of all known substance abuse treatment admissions, regardless of the source of funding for the treatment. The scope of admissions included in TEDS is affected by differences in state reporting practices, varying definitions of treatment admission, availability of public funds, and public funding constraints.

Authorization: In 1988, the Comprehensive Alcohol Abuse, Drug Abuse, and Mental Health Amendments (P.L. 100-690) established a revised Substance Abuse Prevention and Treatment (SAPT) Block Grant and mandated federal data collection on clients receiving treatment for either alcohol or drug abuse. The TEDS data collection effort represents the federal response to this mandate.

Strengths: TEDS contains a large number of AI/AN respondents. The data are collected on a key policy issue, substance abuse. Key demographic indicators are included for each state. One can identify by state, the number of admissions by race, age, gender, and education. There are multiple years of data available, and in addition to the annual files, there is a multi-year file available. Online analysis and subsetting, as well as "Quick Tables" online table generation, are available.

Limitations: Several limitations are identified in the TEDS documentation that should be considered:

- The way an admission is defined may vary from state to state such that the absolute number of admissions is not a valid measure for comparing states.
- The number and client mix of TEDS records depends, to some extent, on external factors, including the availability of public funds. In states with higher funding levels, a larger percentage of the substance-abusing population may be admitted to publicly-funded treatment, including the less severely impaired and the less economically disadvantaged.
- Public funding constraints may direct states to selectively target special populations. For example, pregnant women or adolescents may be more likely to receive treatment. The representations of these populations in the data may vary accordingly.

- States vary in the extent to which coercion plays a role in referral to treatment. This variation derives from criminal justice practices and differing concentrations of abuser subpopulations.
- TEDS consists of treatment admissions, and therefore may include multiple admissions for the same client. Thus, any statistics derived from the data will represent admissions, not clients. It is possible for clients to have multiple initial admissions within a state and even within providers that have multiple treatment sites within the state. TEDS provides a national snapshot of what is seen at admission for treatment, but is currently not designed to follow individual clients through a sequence of treatment episodes.

Access Requirements and Use Restrictions: Data are available to the public at no cost.

Contact Information:

SAMHDA User Support
Substance Abuse and Mental Health Data Archive
Inter-University Consortium for Political and Social Research (ICPSR)
P.O. Box 1248
Ann Arbor, Michigan 48106
(1-888) 741-7242
samhda-support@icpsr.umich.edu
www.icpsr.umich.edu

Data can be accessed at: http://webapp.icpsr.umich.edu/cocoon/SAMHDA-SERIES/00056.xml

Uniform Crime Reports (UCR)

Sponsor: U.S. Department of Justice (DoJ)/Federal Bureau of Investigation (FBI)

Description: The Uniform Crime Reporting (UCR) Program is a nationwide, cooperative summary statistical effort of more than 17,000 city, university and college, county, state, tribal, and federal law enforcement agencies voluntarily reporting data on crimes brought to their attention. The UCR Program collects offense information for murder and nonnegligent manslaughter, forcible rape, robbery, aggravated assault, burglary, larceny-theft, motor vehicle theft, and arson. It also collects information on the characteristics of persons arrested, victims and offenders in homicides and nonnegligent manslaughter, and offenders in hate crimes.

Relevant Policy Issues: Rates of Involvement with Justice System.

Data Type(s): Registry

Unit of Analysis: The unit of analysis is arrests. One person may be arrested multiple times during the year; as a result, the arrest tabulations cannot be considered as a total number of individuals arrested.

Identification of AI/AN/NA: According to the UCR Handbook, revised in 2004, the racial categories used in the UCR Program were adopted from the *Statistical Policy Handbook (1978)* published by the Office of Federal Statistical Policy and Standards, U.S. Department of Commerce. The racial designations are defined as follows:

- White. A person having origins in any of the original peoples of Europe, North Africa, or the Middle East.
- Black. A person having origins in any of the black racial groups of Africa.
- American Indian or Alaskan Native (AI/AN). A person having origins in any of the original peoples of North America and who maintains cultural identification through tribal affiliation or community recognition.
- Asian or Pacific Islander. A person having origins in any of the original peoples of the Far East, Southeast Asia, the Indian subcontinent, or the Pacific Islands. This area includes, for example, China, India, Japan, Korea, the Philippine Islands, and Samoa.

AI/AN/NA Population in Data Set: The total number of arrests of AI/ANs in 2004 was 135,479 for all ages. Of this total, 20,391 AI/AN arrests involved individuals who were under 18 years of age. Total arrests of AI/AN offenders in hate crimes was 41. Information was not available on offenders or victims of homicide because in the published tables, AI/AN is combined with other races into an "other" race category. However, this information is available in the raw data sets, which are available from the UCR Program.

Uniform Crime Reports (UCR)
(continued)

Geographic Scope:	The geographic scope of the reporting system is national. Analyses are presented for principal cities in metropolitan statistical areas (MSAs), metropolitan counties (counties within an MSA), nonmetropolitan counties (counties outside an MSA), and suburban areas (counties within an MSA but excluding principal city). Breakdowns of the data are also available regionally and by population group as well.
Date or Frequency:	Law enforcement agencies submit data on a monthly basis and the data are compiled into annual files. Data are published in annual reports.
Data Collection Methodology:	Law enforcement agencies contribute crime data through their respective state UCR Program. For those states that do not have a state program, local agencies submit crime statistics directly to the FBI.
Participation:	Optional, without incentives
Response Rate:	During 2004, law enforcement agencies active in the UCR Program represented 94.2 percent of the total number of law enforcement agencies.
Strengths:	The data source contains a large AI/AN population. Data are collected on a key policy issue, involvement with the justice system. There are multiple years of data available. This data source is a vast compilation of published tables that are widely used for tracking crime trends across the nation. Additionally, an archive of master files (final data, not estimates) are available upon request.
Limitations:	These are summary data that do not allow analyses beyond simple tabulations by geographic unit, race, and broad age groupings. Moreover, these are primarily tabulations of arrests and, in some cases victim information, so these data cannot be used to determine the number of unique individuals who have been arrested within a year. Some offense data for each year are estimated (arrest data are not estimated) because not all law enforcement agencies are able to provide data for complete reporting periods. The estimates are computed by using the known offense figures of similar areas within a state and assigning the same proportion of crime volumes to nonreporting agencies or agencies with missing data. The estimation process considers the following: population size of agency, type of jurisdiction (e.g., police department versus sheriff's office), and geographic location.
Other:	The National Incident-Based Reporting System (NIBRS), which is compiled in addition to the UCR as summary reporting, covers 80 percent of the nation's reporting. The NIBRS collects data on each single incident and arrest within 22 crime categories. For each offense known to police within these categories, incident, victim, property, offender, and arrestee information are gathered when available. The goal of the redesign is to modernize crime information by collecting data currently maintained in law enforcement records while maintaining the integrity of UCR's long-running statistical series. Implementation of the NIBRS is occurring at a pace commensurate with the resources, abilities,

and limitations of the contributing law enforcement agencies. In, 2004, 29 state programs had been certified for NIBRS participation. For current UCR reporting, NIBRS data are summarized in order to be combined with the UCR data. Contact information for the NIBRS is the same as for the UCR.

Contact Information:

The general website for the UCR is http://www.fbi.gov/ucr/ucr.htm. At this website, reports and statistical tables are available for 1995-2005 (provisional data).

For more information regarding the FBI's UCR data, you may contact a member of the Communications Unit staff by telephone at (304) 625-4995; by facsimile at (304) 625-5394; or by Internet at cjis_comm@leo.gov. (E-mail data requests cannot be processed unless requesters include their full name, a mailing address, and a contact telephone number.)

Reports of Interest:

http://www.fbi.gov/ucr/cius_04/documents/CIUS2004.pdf

http://www.fbi.gov/ucr/cius_04/appendices/appendix_06.html

United States Renal Data System (USRDS)

Sponsor:	U.S. Department of Health and Human Services (DHHS)National Institutes of Health (NIH)/National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)
Description:	The United States Renal Data System (USRDS) is a national data system that collects, analyzes, and distributes information about end-stage renal disease (ESRD) and chronic kidney disease (CKD) in the United States. The USRDS is funded directly by NIDDK in collaboration with the Centers for Medicare & Medicaid Services (CMS). The United Network for Organ Sharing (UNOS) is also providing transplant and wait-list data, under the inter-agency agreement, to this data collection effort in order to improve the accuracy of ESRD patient information.
Relevant Policy Issues:	Measurement of Health Status and Disease-specific Measurements.
Data Type(s):	Registry
Unit of Analysis:	Individual
Identification of AI/AN/NA:	Native American (includes American Indians and Alaska Natives.) The combined category Asian/Pacific Islander is used in USRDS reports.
AI/AN/NA Population in Data Set:	The population size by race is available in Reference Section M of the 2005 USRDS Annual Data Report (ADR). ESRD incidence and prevalence rates by year for Native Americans (NA) are available in Reference Sections A and B. For example, the incident count and adjusted incident rate per million population for NA in 2003 was 1,097 and 503.9 respectively.
Geographic Scope:	The geographic scope of the study is national. Further geographic analysis is possible by state, county, zipcode, and HSA (CDC Health Service Area).
Date or Frequency:	Data have been compiled annually since 1988, with the 2004 data being the most recently available for analysis.
Data Collection Methodology:	Data for the USRDS Database are compiled from existing data sources including the Centers for Medicare and Medicaid Services (CMS) Renal Management Information System (REMIS), CMS claims data, Facility survey data, CDC survey data, Standard Information Management System (SIMS), Medicare Evidence Form (CMS-2728), ESRD Death Notification Form (CMS-2746), and UNOS transplant and wait-list data. The CMS data files are supplemented by CMS with enrollment, payer history, and other administrative data to provide utilization and demographic information on ESRD patients.
Participation:	Mandatory

United States Renal Data System (USRDS)
(continued)

Response Rate: Response or coverage rates are 100 percent since May of 1995 because the amended ESRD entitlement policy requires a Medicare Evidence form to be submitted for all ESRD patients regardless of their insurance and eligibility status. However, the payment data for non-Medicare ESRD patients maybe absent during the 30-month coordination period.

Strengths: Data set contains information on all Native American patients with ESRD. There are multiple years of data available starting from 1978.

Limitations: Payment data of non-Medicare or MSP patients during the first 30-month coordination period is not available.

Access Requirements and Use Restrictions: The data are available to the public through a data use agreement (DUA). The cost associated with use of the data set is $600 for the core data file and $100-$400 for each supplemental file. Researchers completing a data use agreement can access the limited data set directly. To submit a DUA request, contact the Coordinating Center at (888) 99USRDS. Statistical reports providing frequencies and basic tabulations are available through the Renal Data Extraction and Referencing System (RenDER) on the USRDS website: www.usrds.org.

Contact Information: Annual Data Reports can be accessed through:
http://www.usrds.org/adr.htm.

Data requests and publications:
USRDS Coordinating Center
914 South 8th Street
Suite D-206
Minneapolis, MN 55404
(612) 347-7776
(888) 99USRDS
Fax: (612) 347-5878
usrds@usrds.org

Data file contact: Shu Chen, MS, schen@usrds.org.

Washington State Population Survey (WSPS)

Sponsor: State of Washington/Office of Financial Management

Description: The Washington State Population Survey (WSPS) is a source of information about the health and welfare of Washington families. The survey focuses primarily on issues of employment, family poverty, migration into the state, health, and health insurance coverage.

Relevant Policy Issues: Measurement of Health Status, Factors Contributing to Measured Health Disparities, Income Status, Unemployment Rates, Economic Assistance Program Participation Rates, and Economic Opportunity.

Data Type(s): Survey

Unit of Analysis: Individual

Identification of AI/AN/NA: Instructions for reporting race are as follows: “What racial group or groups best describes you?"

- White
- Black
- American Indian or Alaska Native (AI/AN)
- Native Hawaiian or other Pacific Islander (NH/PI)
- Asian

Respondents could select more than one race.

AI/AN/NA Population in Data Set: The 2004 WSPS gathered data on 17,788 individuals from 7,097 households. The unweighted counts for AI/AN respondents to the 2004 WSPS are as follows:
AI/AN alone: 265
AI/AN and other races: 225

Unweighted counts of NH/PI respondents for the 2004 WSPS are as follows:
NH/PI alone: 119
NH/PI and other races: 54

Geographic Scope: The geographic scope of the study is the state of Washington. Geographic areas are identified by regions of the state (e.g., North Puget, West Balance, King, Other Puget Metro, Clark, East Balance, Spokane, Tri-cities). The level of geographical analysis possible is state-wide or regional. However, regional analysis is not recommended for AI/AN/NA subgroups because of very small sample sizes.

Date or Frequency: The WSPS has been conducted biennially since 1998. The 2006 administration of the survey was underway at the time of preparation of this catalog.

Washington State Population Survey (WSPS)
(continued)

Data Collection Methodology:	The WSPS is a telephone survey.
Participation:	Optional, without incentives
Response Rate:	WSPS survey contacts indicated that the response rate for the 2004 study was low.
Sampling Methodology:	For the WSPS, a stratified sample by region was selected from all households in state of Washington with an activated telephone line, either listed or unlisted. A target of 750 households was planned for each of the eight regions with the exception of King County where a target of 1,800 households was planned. More households were selected in King County to insure sufficient information on racial minority groups for statistical analyses. Households were selected in each of the regions using random digit dialing (RDD) technique.
Analysis:	There are two weights for use in analysis. A population weight is available that weights the survey responses to represent the state population based on Census Bureau population counts. There is also a weight based on administrative records for Medicaid that, when used, will yield improved estimates for uninsured persons.
Strengths:	Data are collected on key policy issues including health and economic status. There are multiple years of data available.
Limitations:	There are limited AI/AN/NA respondents in this data source. This study has a low response rate.
Access Requirements and Use Restrictions:	Data are available to the public at no cost.
Contact Information:	The data and documentation are available for download from the following website: http://www.ofm.wa.gov/sps/2004/default.asp. Erica Gardner Forecasting Division Washington State Office of Financial Management P.O. Box 43113 Olympia, WA 98504-3113 (360) 902-0599

Youth Gangs in Indian Country

Sponsor:	U.S. Department of Justice (DOJ)/Office of Juvenile Justice and Delinquency Prevention (OJJDP) and the National Youth Gang Center (NYGC)
Description:	In 2001, OJJP and NYGC developed and implemented the 2000 Survey of Youth Gangs in Indian Country. All federally recognized Indian communities were surveyed to measure the presence, size, and criminal behavior of youth gangs in Indian Country. This survey collected data regarding the presence and effect of youth gang activity in Indian Country as well as programmatic responses to the problem. The survey was mailed to tribal leaders or tribal representatives in 577 Indian communities comprising 561 federally recognized tribes.
Relevant Policy Issues:	Rates of Involvement with Justice System.
Data Type(s):	Survey
Unit of Analysis:	Federally recognized Indian communities
Identification of AI/AN/NA:	AI/AN/NA individuals are not identified in the data set. Instead, data are organized by Indian community. The survey defines an Indian "community" as persons of American Indian, Alaska Native, or Aleut heritage who reside within the limits of Indian reservations, pueblos, rancherias, villages, dependent Indian communities, or Indian allotments, and who together comprise a federally recognized tribe or community. Communities also include people who have been recognized by the United States government as a tribe or tribal community, but who do not occupy tribal trust, tribally owned, or Indian allotment lands. Communities are the people and land together or tribal community viewed as a group. Land without the people is not considered a community for the purpose of this survey. The data source does not allow identification of members of a state recognized tribe.
AI/AN/NA Population in Data Set:	Race data on the respondents was not gathered in the survey. Communities that reported gang activity in 2000, however, were asked to estimate demographic characteristics of gang members, including race or ethnicity. Survey respondents reported that the majority (78 percent) of youth gang members in their communities were American Indian, Alaska Native, or Aleut. In fact, approximately one-half of responding communities indicated almost all gang members (more than 90 percent) were of this race.
Geographic Scope:	The geographic scope of the study is national. Information on additional geographic indicators was not available.
Date or Frequency:	This study was a one-time effort. Data were collected in 2001.

Youth Gangs in Indian Country
(continued)

Data Collection Methodology:	Mail and telephone survey
Response Rate:	Overall, 52 percent (n=300) of the communities responded to the survey.
Sampling Methodology:	At the time the survey was developed, there were 577 Indian communities in the United States, comprising 561 federally recognized tribes. NYGC and the advisory group chose to survey the entire Indian country population to provide a broad assessment.
Strengths:	Data are collected on the key policy issue of justice system involvement. For those interested in youth gang activity, this data set is the only assessment of such activity across all Indian Country.
Limitations:	The data are very difficult to obtain and the documentation available online does not include information on statistical topics such as error estimates or weighting. There is a low response rate.
Access Requirements and Use Restrictions:	The data set, held by the NYGC, is only available on a very limited basis. It is not intended for release, but release has occurred in a few exceptional situations. There is no cost should access be granted.
Contact Information:	National Youth Gang Center Institute for Intergovernmental Research Post Office Box 12729 Tallahassee, FL 32317 Tel: (850) 385-0600 Fax: (850) 386-5356

Youth Risk Behavior Surveillance System (YRBSS)

Sponsor:	U.S. Department of Health and Human Services (DHHS)/Centers for Disease Control and Prevention (CDC)
Description:	The Youth Risk Behavior Surveillance System (YRBSS) is an epidemiologic survey system established by CDC to monitor the prevalence of youth behavior that most influences health. The priority health risk behaviors that contribute markedly to the leading causes of death, disability, and social problems among youth and adults in the United States include tobacco use; unhealthy dietary behaviors; inadequate physical activity; alcohol and other drug use; sexual behaviors that contribute to unintended pregnancy and sexually transmitted diseases including HIV infection; and behaviors that contribute to unintentional injuries and violence.
Relevant Policy Issues:	Measurement of Health Status, Key Health Disparities, and Factors that Contribute to Health Disparities.
Data Type(s):	Survey
Unit of Analysis:	Individual
Identification of AI/AN/NA:	Race/ethnicity are ascertained by the following two questions: Are you Hispanic or Latino? • Yes • No What is your race? (Select one or more responses.) • American Indian or Alaska Native (AI/AN) • Asian • Black or African American • Native Hawaiian or other Pacific Islander (NH/PI) • White
AI/AN/NA Population in Data Set:	For 2005, there are 13,917 records for the National Youth Risk Survey; of these 147 are identified as AI/AN and 90 as NH/PI.
Geographic Scope:	The geographic scope of the study is national. Geographic identifiers available for analysis include geographic region (Northeast, Midwest, South, or West). Prior to 2005, metropolitan status (urban, suburban, or rural) is available, but will not be made available in data sets beginning in 2005.
Date or Frequency:	School-based data have been collected in odd years since 1991. The 2005 National School-based Youth Risk Behavior Survey is available for public use.

National Alternative High School YRBSS was conducted in 1998 and the National College Health Risk Behavior Survey was conducted in 1995.

Data Collection Methodology: Students complete the self-administered questionnaire in their classrooms during a regular class period, and record their responses directly on a computer-scannable booklet or answer sheet.

Participation: Optional, without incentives to students.

Response Rate: In 2005, the school response rate was 78 percent and the student response rate was 86 percent. When these response rates are combined, the overall response rate equaled 67 percent.

Sampling Methodology: The 2005 national school-based survey employed a three-stage cluster sample designed to produce a nationally representative sample of students in grades 9-12. The first stage sampling frame contained primary sampling units (PSUs) consisting of large counties, sub-areas of very large counties, or groups of small, adjacent counties. The PSUs were selected with probability proportional to school enrollment size. At the second sampling stage, 195 schools were also selected with probability proportional to school enrollment size. To enable separate analysis of data for black and Hispanic students, schools with substantial numbers of black and Hispanic students were sampled at higher rates than all other schools. The third stage of sampling consisted of randomly selecting one or two intact classes of a required subject (e.g., English or social studies) from grades 9-12 at each chosen school. All students in the selected classes were eligible to participate in the survey.

Strengths: Data are collected on key policy issues, including health status, health disparities, and factors that contribute to key health disparities. Multiple years of data are available for trend analysis.

Limitations: There are few AI/AN/NA respondents in each year. These low numbers will make complex analyses on these populations difficult.

Other: The Bureau of Indian Affairs (BIA) and the Navajo Nation also conduct the YRBS on about a 3-year cycle. These data are owned by the BIA and the Navajo Nation. Potential users can contact BIA and the Navajo Nation for information about accessing these data.

For access to the YRBSS data for the Navajo Nation, it is probable that the Nation will need to approve the use of data by outside researchers through the Navajo Nation Health and Human Research Review Board. The proposing party would be required to submit a proposal to the Navajo Nation with supporting documents on the purpose and use of the data, and benefits of outcome for the Navajos. A report summarizing the results of the Navajo Nation's YRBSS in 1997, 2000, and 2003 is available for public dissemination at the following website: www.yrbs.navajo.org.

Youth Risk Behavior Surveillance System (YRBSS)
(continued)

Contact information concerning these data follows:

BIA:
Jack Edmo at JEdmo@bia.edu or (505) 248-6964

Navajo Nation:
Christine J. Benally, Ph.D.
Lead Epidemiologist, Community Health Services
CDR USPHS, Director Support Scientist Officer
P. O. Box 160, N. U.S. Hwy 491
Shiprock, NM 87420
(505) 368-7427 desk
(505) 368-6324 fax
(505) 368-6300 office
christine.benally@ihs.gov@

Access Requirements and Use Restrictions:

YRBSS data are available to the public at no cost.

Contact Information:

Dr. Laura Kann
Division of Adolescent and School Health
National Center for Chronic Disease Prevention and Health Promotion
Centers for Disease Control and Prevention
Mailstop K-33
4770 Buford Highway, NE
Atlanta, GA 30341-3717
(770) 488-6181
LKK1@cdc.gov

Or: healthyyouth@cdc.gov
(770) 488-6161

Data can be accessed at www.cdc.gov/yrbs.

6. Other Data Sources

The following list describes data sources that were considered but not profiled for the catalog. Reasons for excluding data sources include the inability to identify AI/AN/NA individuals in the data source; a very limited number of AI/AN/NA individuals in the data source; lack of relevance to the identified key policy issues; extended length of time since the study was conducted; or data unavailable in any form for independent research. This list describes some basic information about these data sources and includes the reason(s) for excluding each data set.

Alcohol and Drug Services Study

This study involves a drug treatment facility and client sample survey, in which data are collected to estimate the length of a patient's stay, the cost of treatment, and to describe post-treatment status of clients. This study is excluded from the catalog as there are only 59 AI/AN individuals identified in the data set.

Additional information about this study can be found at:
http://www.oas.samhsa.gov/ADSS/ADSS2ClientCB.pdf

Drug Abuse Treatment Outcome Study: Adolescent (DATOS-A)

The study was designed to determine the outcomes of drug abuse treatment delivered in typical community-based programs by examining the role of treatment outcomes and program type, client characteristics, treatment received, therapeutic approaches, and provision of aftercare. Information was collected on a small number of AI/AN individuals, but these individuals cannot be identified in the data set.

For more information about this study, contact:
SAMHDA/ICPSR
The University of Michigan
P.O. Box 1248
Ann Arbor, MI 48106-1248
SAMHDA Helpline: 1-888-741-7242
e-mail: samhda-support@icpsr.umich.edu

Drug and Alcohol Services Information System (DASIS)

This system is part of the National Directory of Drug and Alcohol Abuse Treatment Programs and the online Substance Abuse Treatment Facility Locator. It was excluded from this catalog because of its focus on facilities rather than individuals or families.

Additional information about the DASIS can be found at:
http://oas.samhsa.gov/dasis.htm#DASISinfo

Evaluation of the Tribal Strategies Against Violence (TSAV) Initiative in Four Tribal Sites in the United States, 1995

This study was designed to develop comprehensive strategies in tribal communities to reduce crime, violence, and substance abuse. Approximately 90 interviews were conducted in four locations. Due to the limited number of sites and respondents and extremely limited access to the data, this study is excluded from the catalog.

Questions regarding this study can be addressed to:
Director
National Archive of Criminal Justice Data
Inter-University Consortium for Political and Social Research
Institute for Social Research
P.O. Box 1248
University of Michigan
Ann Arbor, MI 48106-1248

Family Data on Public and Indian Housing

This study of public and Indian housing projects and their tenants is 13 years old. The records are summaries for housing projects that do not provide individual-level information. Also, there is little documentation.

Questions regarding this study can be addressed to:
Information Services Division of Public and Indian Housing
(202) 708-1445

Gambling Impact and Behavior Study, 1997-1999

This study on the gambling behavior of American adults and youth and the impact of gambling facilities on local economies is excluded because it does not identify people who are AI/AN/NA, nor does it focus on a specific geographic region with a large AI/AN/NA population.

The data, codebooks, and other documentation are available at: http://www.icpsr.umich.edu/cgi-bin/bob/newark?study=2778&path=SAMHDA

Hawaii Student Alcohol, Tobacco and other Drug Use Survey

The purpose of this study is to assess the incidence and prevalence of alcohol, tobacco, and other drug use among students in grades 6 through 12 throughout the State of Hawaii. Results are based on student responses from public, private, and charter schools. The 2003 data set contains answers from 4,912 Native Hawaiian students. However, the Alcohol and Drug Abuse Division (ADAD) of the Hawaii Department of Health does not release this data for use by researchers.

Additional information about this study can be found at:
Alcohol and Drug Abuse Division
601 Kamokila Blvd.
Kapolei, Hawai‘i, 96707
(808) 692-7506
http://www.hawaii.gov/health/substance-abuse/prevention-treatment/survey/adsurv.htm

Hawaiian Community Survey

This is an annual survey of households in the state of Hawaii in which at least one member is identified as Hawaiian. The survey was conducted from 2001 to 2005 as part of an effort to assess the true educational needs in Hawaii. The survey contains information on family well-being, childcare arrangements for preschool-age children, obstacles and achievements among school-age children, and continuing educational pathways among adults. At this time, there is no public use data file available for researchers, although the data may become available in the future.

More information about this study can be found at:
http://www.ksbe.edu/pase/researchproj-hicomsrvy.php

Head Start Family and Child Experiences Survey (FACES)

This study provides longitudinal information on a periodic basis on the characteristics, experiences, and outcomes for children and families served by Head Start. The study does not include American Indian/Alaska Native Head Start Programs and does not identify AI/AN/NA as one or more separate race categories.

More information about this study can be found at:
http://www.acf.hhs.gov/programs/opre/hs/faces/index.html

Health and Diet Survey

This periodic telephone survey measures and monitors public awareness, knowledge, attitudes, and reported behavior related to food and nutrition. Although the survey includes a race category for American Indians/Alaskan Natives, the number in this category is so small (n<100) that it is combined with other small categories into “other” for the public use data set. The survey also does not lend itself to analysis of a specific geographic region relevant to the AI/AN/NA population because of its small size (N=1,798).

Additional information regarding this study can be found at:
http://www.cfsan.fda.gov/~comm/crnutri3.html#demog

Medicare Current Beneficiary Survey

This survey of Medicare beneficiaries residing in the United States and Puerto Rico contains very few AI/AN/NA cases in the sample (n=36 for the last round of data collection) and aggregation across years is not recommended.

Additional information on this study can be found at: http://www.cms.hhs.gov/apps/mcbs/

Monitoring the Future: A continuing study of American youth

This ongoing study of the behaviors, attitudes, and values of American youths contains very few American Indian (AI) respondents (i.e., approximately 1 percent of the grade 12 samples). For this reason, the public use data sets and restricted-access data sets do not contain an indicator for AI in the data set.

Additional information regarding this study can be found at:
http://www.monitoringthefuture.org/

National Health and Nutrition Examination Survey (NHANES)

This survey collects data on the health and nutritional status of children and adults in the U.S. Because there is no oversample for Native American/Alaskan Natives in the NHANES, the sample of AI/AN/NA is very small and is grouped into the "other" category in the public use files.

Additional information regarding this study can be found at:
http://www.cdc.gov/nchs/nhanes.htm

National Immunization Survey (NIS)

Data from the NIS are used to produce timely estimates of vaccination coverage rates. The data do not identify the AI/AN/NA populations of interest for this catalog, as the public-use data files do not identify AI/AN/NA individuals.

Additional information regarding this study can be found at:
http://www.cdc.gov/nis/

National Longitudinal Survey of Youth

This study focuses on the labor market experiences of American youth. The study included an oversample of blacks, Hispanics, and disadvantaged whites. However, the sample of AI/AN/NA individuals is too small to identify as a separate racial category in the data set.

Additional information regarding this study can be found at:
http://www.bls.gov/nls/nlsy97.htm

National Survey of Substance Abuse Treatment Services

This study is a compilation of data collected from facilities and is used to update the National Directory of Drug and Alcohol Abuse Treatment Programs and the online Substance Abuse Treatment Facility Locator. This study was excluded from this catalog because its focus on facilities rather than individuals or families.

Additional information regarding this study can be found at:
http://oas.samhsa.gov/dasis.htm#nssats2

Native Hawaiian Children and Families - Provider Survey and Consumer Survey

The goal of the Consumer Survey was to gather information on consumers' perceptions of the services they receive from programs located in their communities, while the goal of the Service Provider Survey was to gather information from community agencies about their programs and the services they provide to Native Hawaiian children and families. It could not be ascertained whether the data from this study are publicly available.

Additional information regarding this study can be found at:
http://uhfamilydata.hawaii.edu/hi_child_ed/hi_child_ed.asp

Navajo Health and Nutrition Survey

The Navajo Health and Nutrition Survey was conducted from 1991 to 1992 to assess the health and nutritional status of Navajo Reservation residents using a population-based sample. The data do not appear to be available in any format.

Property Owners and Managers Survey (POMS)

The Property Owners and Managers Survey (POMS) was designed to learn more about rental housing and the providers of rental housing. A nationwide sample of approximately 16,300 housing units that were rented or vacant-for-rent in the 1993 American Housing Survey National Sample (AHS-N) was selected, and a questionnaire was mailed to the property owner, manager, or other agent of the owner of each property containing a selected unit. This study is excluded because there are very few AI/AN/NA respondents in the data set and the data cannot be combined with other data to increase the number of AI/AN/NAs. In addition, the survey content focuses on the characteristics of the rental properties that the respondents own or manage, not the personal dwellings of the respondents.

Contact the Financial and Market Characteristics Branch at (301) 763-3199 or visit ask.census.gov for further information on Property Owners & Managers Survey (POMS) Data. The actual data are hosted on-line by HUDUSER and are available for downloading from the following site: http://www.huduser.org/datasets/poms.html

Social Security Administration Benefits and Earnings Public Use File

The 2004 Benefits and Earnings Public-Use File consists of a 1 percent random, representative sample of records of Old-Age, Survivors, and Disability Insurance beneficiaries who were entitled to receive a Social Security (OASDI) benefit for December 2004, and all benefit information is as of December 2004. This file does not contain any racial indicators.

More information about this study can be found at:
http://www.socialsecurity.gov/policy/docs/microdata/earn/index.html

Study of Tribal and Alaska Native Juvenile Justice Systems in the United States, 1990

This congressionally-mandated analysis of tribal juvenile justice systems was conducted in 1990-1992. We could not find any indication that these data are available to researchers.

Suicide and Risk Behaviors in an Incarcerated American Indian Population in the Northern Plains [United States], 1999-2000

This multi-part study was an evaluation of five intake screening protocols in a county jail in the Northern Plains. The study is being excluded from the catalog because it was designed only to evaluate and compare the five protocols and the data collected cannot be combined to facilitate other analyses.

Questions regarding this study can be addressed to:
Director
National Archive of Criminal Justice Data
Inter-university Consortium for Political and Social Research
Institute for Social Research
P.O. Box 1248
University of Michigan
Ann Arbor, MI 48106-1248

Survey of Active Duty Personnel

This study focuses on the experiences, attitudes, and demographic characteristics of all Army, Navy, Marine Corps, Air Force, and Coast Guard active-duty members. Although there is a public use data file available for this study, AI/AN individuals can not be identified and geographic analysis of regions specific to this population is not possible.

For more information concerning this study, contact Dr. Jim Caplan at 703-696-5848.

Survey of American Indians and Alaska Natives (SAIAN)

The National Medical Expenditure Survey (NMES) series provides information on health expenditures by or on behalf of families and individuals, the financing of these expenditures, and each person's use of services. Conducted in 1987, the Survey of American Indians and Alaska Natives (SAIAN) was designed in collaboration with the Indian Health Service (IHS), and used the same data collection instruments, interview procedures, and time frame as the NMES Household Survey component. However, the SAIAN differed from the Household Survey in several respects. The SAIAN sample was interviewed only three times and was not given the supplements on long-term care, caregiving, and care-receiving. Also, SAIAN respondents were asked additional questions on topics such as use of IHS facilities and traditional medicine, and were given a modified self-administered questionnaire with separate versions for adults and children. Interviewers for the SAIAN were mainly American Indians or Alaska Natives, and about 20 percent of the interviews were not conducted entirely in English. Of these, approximately 40 percent were conducted entirely in the native language of the respondent. Data were collected on 7,071 AI/AN persons in the eligible dwelling units.

Although the topic of this study is of clear relevance to the AI/AN population, this study is excluded from this catalog as these data are nearly 20 years old and cannot be combined with other data.

For more information about this study, contact ICPSR User Support at netmail@icpsr.umich.edu or call (734) 647-2200 for information about accessing ICPSR data.

Survey of Health Related Behaviors among Military Personnel

This study provides comprehensive and detailed estimates of the prevalence of alcohol, illicit drug, and tobacco use and the negative effects of alcohol use among active-duty personnel. Although there is a public use data file available for this study, AI/AN individuals can not be identified and geographic analysis of regions specific to this population is not possible.

Questions regarding this study can be addressed to:
LTC Lorraine Babeu, Ph. D., CCC-A
Office of the Assistant Secretary of Defense (Health Affairs)
Tricare Management Activity
Health Program Analysis & Evaluation
5111 Leesburg Pike, Skyline 5, Suite 510
Falls Church, VA 22041-3206
Office: (703) 681-3636 DSN: 761-3636
Fax: (703) 681-3682

Survey of Income and Program Participation (SIPP)

The purpose of this study is to collect source and amount of income, labor force information, program participation and eligibility data, and general demographic characteristics to measure the effectiveness of existing federal, state, and local programs; to estimate future costs and coverage for government programs, such as food stamps; and to provide improved statistics on the distribution of income and measures of economic well-being in the country. However, the data files use the following race categories: White alone, Black alone, Asian alone, and Other. There is no way to identify AI/AN/NA individuals in the data set.

More information about SIPP can be found at: http://www.bls.census.gov/sipp/

7. Other Reports

During the course of constructing the data catalog, the project team identified a number of useful reports that are listed below. This section is not meant to serve as a comprehensive listing of reports relevant to the topic of this catalog, but rather to serve as a potentially useful resource for users of the catalog.

1. **American Indian and Alaska NativeChildren: Findings From the Base Year of the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B).** 2005. Flanagan, K., and Park, J. (NCES 2005–116). U.S. Department of Education. Washington, DC: National Center for Education Statistics

2. **American Indian and Alaska Native Roundtable on Long Term Care: Final Report 2002.** Roundtable Report Prepared by: Jo Ann Kauffman, Kauffman and Associates, Incorporated, 425 West First Avenue, Spokane, WA.

3. **American Indians and Crime.** Lawrence A. Greenfeld and Steven K. Smith, BJS Statisticians. February 1999, NCJ 173386 http://www.ojp.usdoj.gov/bjs/abstract/aic.htm

4. **American Indian Reservations: Montana, North Dakota, and South Dakota.** Pilot Project. Specialty Products, Part 1. AC-02-SP-1. October 2004. U.S. Department of Agriculture. http://www.nass.usda.gov/census/amindian.pdf

5. **American Indians on Reservations: A Databook of Socioeconomic Change between the 1990 and 2000 Census.** Jonathan B. Taylor and Joseph P. Kalt. January 2005. The Harvard Project on American Indian Economic Development. http://www.ksg.harvard.edu/hpaied/pubs/pub_151.htm

6. **Background Report on the Use and Impact of Food Assistance Programs on Indian Reservations.** K. Finegold, N. Pindus, L. Wherry, S. Nelson, T. Triplett, R. Capps. January 2005. The Urban Institute.

7. **Characteristics of Food Stamp Households: Fiscal Year 2004.** Anni Poikolainen. September 2005. Mathematica Policy Research, Inc.

8. **Family Violence and American Indians/Alaska Natives: A Report to the Indian Health Service Office of Women's Health.** Principal Investigator: Laura A. Williams. http://www.ihs.gov/PublicInfo/PublicAffairs/PressReleases/Press_Release_2002/Compendium_Part_I_and_II.pdf

9. **HIV/AIDS Surveillance Report, 2004.** Centers for Disease Control and Prevention. Vol. 16. Atlanta: US Department of Health and Human Services, Centers for Disease Control and Prevention; 2005. http://www.cdc.gov/hiv/stats/hasrlink.htm.

10. **Indian Health Service Population Estimates and Projections** http://www.ihs.gov/NonMedicalPrograms/IHS_Stats/Statistical_Databases.asp

11. **Ka huaka'i: 2005 Native Hawaiian Educational Assessment.** S.K. Kana'iaupuni, N. Malone, and K. Ishibashi. Honolulu, HI: Kamehameha Schools, Pauahi Publications.

12. **Longitudinal Study of the Indian Vocational Rehabilitation Services Program.** May 2003. Final Report was submitted to Rehabilitation Services Administration, U.S. Department of Education, in partial fulfillment of requirements under ED Contract No. HR92022001. http://www.ilr.cornell.edu/ped/lsvrsp/PublishedResearchFiles/RTI_1stFINAL_Report.pdf

13. **Office of Indian Education Programs Annual School Report Card** http://www.oiep.bia.edu/Report%20Cards/Annual%20Report%20Card%2004-05.htm

14. **Racial and Ethnic Disparities in the Experiences of Health Care Consumers.** Karen Onstad. A research brief published in November 2005 by the National Consumer Assessment of Health Care Providers and Systems (CAHPS) Benchmarking Database, under AHRQ Contract Number 290-0I-0003.

15. **Regional Differences in Indian Health.** Rockville, MD: Indian Health Service, 1998-99. http://www.ihs.gov/PublicInfo/Publications/trends98/region98.asp

16. **Sexually Transmitted Disease Surveillance, 2004.** Centers for Disease Control and Prevention. Atlanta, GA: U.S. Department of Health and Human Services, September 2005.

17. **State Tobacco Activities Tracking and Evaluation (STATE) System.** Centers for Disease Control and Prevention. Available at: http://www.cdc.gov/tobacco/statesystem.

18. **Status and Trends in the Education of American Indians and Alaska Natives.** C. Freeman and M. Fox. (2005). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office, NCES 2005-108.

19. **Strong Heart Study Data Book: A Report to American Indian Communities.** November 2001. National Institutes of Health, National Heart, Lungs and Blood Institute, Division of Epidemiology and Clinical Applications. NIH Publication No. 01-3285. http:www.nhlbi.nih.gov/resources/docs/shs_db.htm

20. **Summary of Notifiable Diseases – United States, 2004.** Centers for Disease Control and Prevention. Published June 16, 2006, for MMWR 2004: 53 (No. 53). http://www.cdc.gov/mmwr/summary.html

21. **Tobacco Use Among U.S. Racial/Ethnic Minority Groups—African Americans, American Indians and Alaska Natives, Asian Americans and Pacific Islanders, and Hispanics: A Report of the Surgeon General.** U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. 1998http://www.cdc.gov/TOBACCO/sgr/sgr_1998/index.htm

22. **Trends in Indian Health, 2000-2001.** Rockville, MD: Indian Health Service, 2000. http://www.ihs.gov/NonMedicalPrograms/IHS_Stats/Trends00.asp

Appendices

Appendix A: Glossary

Term	Definition
Census Division	Census divisions are groupings of states and the District of Columbia that are subdivisions of the four census regions. There are nine census divisions, which the U.S. Census Bureau established in 1910 for the presentation of census data.
Census Region	Census regions are grouping of states and the District of Columbia. The four regions (Northeast, Midwest, South and West) represent areas that were relatively homogeneous when they were established in 1910 and revised in 1950.
Computer Assisted Personal Interviewing (CAPI)	A method of data collection where the interviewer and respondent conduct an interview face-to-face. The survey questions are displayed on the screen of a computer. The interviewer reads the questions to the respondent and then enters the respondent's answers into a database on the computer.
Computer Assisted Telephone Interviewing (CATI)	A method of data collection where the interviewer and respondent conduct an interview over the telephone. The survey questions are displayed on the screen of a computer. The interviewer reads the questions to the respondent and then enters the respondent's answers into a database on the computer.
Consolidated Metropolitan Statistical Area (CMSA)	A metropolitan area is identified as a CMSA when it has a population of one million or more and also has separate component areas [Primary Metropolitan Statistical Areas (PMSAs)] that meet statistical criteria and are supported by local opinion.
Core Based Statistical Area (CBSA)	The term core based statistical area (CBSA) became effective in 2000 and refers collectively to metropolitan and micropolitan statistical areas. The 2000 standards provide that each CBSA must contain at least one urban area of with a population of 10,000 or more.
Cluster sampling	When the basic sampling unit in the population is to be found in groups or clusters (e.g., individual respondents in households), the sampling is sometimes carried out by first selecting a sample of clusters (e.g., households) and then collecting data from all or a sample of the members of each selected cluster.
Equal Probability Sample	An equal probability sample is derived from a random sampling method in which all units on the sampling frame have an equal probability of being selected.
Design Effect	The design effect is the ratio of the actual variance of a sample to the variance of a simple random sample with the same number of elements. The size of the design effect helps researchers understand the extent to which the data collection approach affects the results.

Term	Definition
Disproportionate Stratified Sample	A disproportionate stratified sample is a sampling method where particular units within a stratum have a higher probability of selection than other units. This method is often used when a sample design calls for an oversample of a particular group (i.e., AI/AN population).
Federal Information Procession Standards (FIPS) code	FIPS codes are a standardized set of numeric or alphabetic codes issued by the National Institute of Standards and Technology (NIST) to ensure uniform identification of geographic entities through all federal government agencies. The entities covered include states and statistically equivalent entities, counties and statistically equivalent entities, named populated and related location entities (such as, places and county subdivisions), and American Indian and Alaska Native areas.
Field Period	The field period denotes the start and end date of a data collection effort.
Metropolitan Statistical Area (MSA)	MSAs are metropolitan areas (MAs) that are not closely associated with other MAs. These areas typically are surrounded by nonmetropolitan counties (or county subdivisions in New England).
Mitofsky-Waksberg sample design	The Mitofsky-Waksberg sample design is a multi-stage sampling design method developed to maximize the efficiency of random digit dialing. Telephone numbers that are assigned to residential households tend to be grouped in banks of consecutive telephone numbers. The Mitofsky-Waksberg method groups telephone numbers into banks of 100 consecutive numbers using the area code, the 3-digit prefix, and the first 2 digits of the suffix. In the first stage of sampling, telephone numbers within the 100-banks are selected at random with replacement to determine if the number comes from a bank of residential numbers. If the telephone number is residential, the 100-bank is retained for the second stage of sampling.
Multi-stage sample	A multi-stage sample is derived from a sampling method that uses more than one stage of selection to draw the sample. The first stage of selection involves selecting the largest sampling unit [also called the primary sampling unit (PSU)], while the final stage of sampling involves selecting the smallest sample unit (for example, an individual or household).
Panel survey/study	A panel survey or study is a type of longitudinal study in which a group of subjects are surveyed on more than one occasion.
Primary Metropolitan Statistical Area (PMSA)	Each Primary Metropolitan Statistical Area (PMSA) consists of a large urbanized county or cluster of counties (or cities and towns in New England) that demonstrate very strong internal economic and social links, in addition to close ties to other portions of the larger area. If an area that qualifies as a metropolitan area (MA) has 1 million people or more, two or more PMSAs may be defined within it.
Primary Sampling Unit (PSU)	A primary sampling unit (PSU) is the first sample entity drawn in a multi-stage sample (in a single stage sample, the primary sampling unit is the first and only sampling unit drawn).

Term	Definition
Random Digit Dialing (RDD)	Random Digit Dialing (RDD) is an automated method of dialing telephone numbers at random in a targeted area. Unlisted telephone numbers are included in the sample frame.
Simple Random Sampling (SRS)	Simple Random Sampling (SRS) is a method of sampling that gives all units in the sample frame an equal chance of being selected for inclusion in the sample, and an equal chance for selection for each of all possible samples of the same size.
Standard error	The standard error is a measure of the variation that might be expected to occur merely by chance in the characteristics of samples drawn equally randomly from the same population. It is the standard deviation of the sampling distribution of a statistic. The smaller the standard error, the better the sample statistic is as an estimate of the population parameter.
Stratified Sample	In stratified sampling, an independent sample is selected from each strata of the population. Strata are divisions of the populations into relevant subsets (e.g., race, location).
Two-stage sampling	A two-stage sample is a simple case of a multi-stage sample. The population to be sampled is first classified into primary sampling units (PSUs) (e.g., census tracts) each of which consists of a number of the basic sampling units or the secondary sampling units (SSUs) (e.g., households within the census tract). A sample of these PSUs is taken, constituting the first stage, and then a sample of the SSUs is taken, constituting the second stage.
Wave	Wave is a term used to describe a single field period for a study that involves collecting data more than once (such as in a panel study).

Sources

Kish, L. *Survey Sampling*. 1965. John Wiley & Sons, Inc., New York, NY.

Pocket Dictionary of Statistics. 2002. The McGraw-Hill Companies, Inc., New York, NY.

U.S. Census Bureau online glossary: http://www.census.gov/main/www/glossary.html

OECD Glossary of Statistical Terms: http://stats.oecd.org/glossary/index.htm

Appendix B: Data Set Aggregation

Researchers may want to combine survey data from two or more data collection efforts (e.g., combining data across years for a particular study) in order to construct an analytic data set containing more respondents than the individual data sets themselves. This combined data set could have greater statistical power and precision to answer research questions. As an example, consider two data sets, one with 400 respondents and the other with 600 respondents. The confidence interval for an estimate of 40 percent is ± 4.8 percent for the smaller survey and ± 3.9 percent for the larger survey.[9] If the surveys are combined, yielding a total of 1,000 respondents, the confidence interval drops to ± 3.0 percent. The combined confidence interval is 38 percent less than that for the smaller survey and 23 percent less than that for the larger survey. Combining the surveys has substantially increased the precision of survey estimates over those for each of the surveys individually. Combining data from multiple surveys is often considered when the research focus is upon relatively rare or under-sampled populations.

Many of the issues involved in determining if survey data can be combined and how they should be combined are substantive and require consideration by subject-matter experts rather than statistical consultation. Moreover, these substantive issues should be resolved before any statistical consultation can take place. This appendix focuses primarily on the substantive issues that need to be addressed in order to decide if two or more data sets can be successfully aggregated. The appendix only briefly touches on the statistical operations that may be required to aggregate multiple data sets. Further analytical and statistical consulting may be required before such an aggregation can be accomplished correctly.

Combining Survey Data Sets: Substantive Questions to Address

The following four questions can serve as a guide to determining the advisability of combining data from multiple survey efforts. Each of these questions is discussed in more detail in the paragraphs that follow:

1. Are the populations the same for the different data collection efforts?
2. Do survey questions and response categories match?
3. Might differences in survey administration dates affect survey results?
4. What were the survey sample designs?

[9] In this example we assume simple random sampling and are using the formula $\pm 1.96*\sqrt{((p*q)/N)}$ for calculating the confidence interval of a proportion.

Are the populations the same for the different data collection efforts? First, we must determine if the surveyed populations match. Are there age, geographic, or other definitions that differ among the surveys? If one survey defines its adult population as 18 to 65 years-of-age and the other as 16 to 85 years-of-age, the differences between the populations may make the meaning of responses somewhat different. If this is so, combining the data sets would not be reasonable. Similarly, the geographic areas sampled may differ between surveys. For example, does it make sense to combine Oregon respondents with those from Wyoming? The determination of matching populations is essentially a substantive, not statistical, issue. The researcher must determine if the differences between the populations will make a substantive difference in the meaning of their responses or if some reconciliation could be made that would correct for those differences in response. In the case of the age difference between the two survey populations presented earlier, the reconciliation may be to either restrict the age range to the most limited or expand it to the most inclusive. In either case, the researcher should be able to clearly justify his/her rationale for deciding to combine the two populations or changing one population to match another should he/she decide to combine data sets where the populations differ.

Do survey questions and response categories match? If the survey populations (frames) are considered comparable, the next question concerns the survey content and question presentation. Here the survey questions of research interest, the response and reporting categories offered, and the context of the questions must be compared. Is the wording of questions of interest in the surveys exactly the same? Or, if not exactly the same, are the wordings close enough to be considered the same? Are the response categories the same or can they be recoded into the same categories? If response categories cannot be altered to match, combining the data is not feasible.

A similar review must be made of the variables that will be used to form subgroups for reporting purposes. Often these reporting subgroups are defined by demographic characteristics such as age, race/ethnicity, or geographic location. Are the subgroups that can be formed for the surveys comparable? Again, if comparable subgroups cannot be formed for the reporting of survey data, it makes no sense to combine survey data.

Finally, does the placement of the questions of interest in the respective questionnaires differ? If so, could this difference influence responses? Stated slightly differently, does the placement of questions in the flow of the questionnaires predispose respondents to respond differently? As with the

judgments that must be made regarding population comparability and survey content difference, this decision will be based on the judgment of the researcher and should be carefully documented.

Might differences in survey administration dates affect survey results? Combining surveys with different administration dates assumes that survey results are not time-dependent. This assumption presumes that survey results for an earlier survey and a later survey are independent and do not differ in a way that relates to the date of survey administration (i.e., no intervening events or changes in circumstances occurred that may influence survey responses). The question for the researcher is whether this assumption is appropriate. As was the case with the previous issues, determination of whether it is reasonable to combine surveys given the time difference between their administrations is an issue that should be carefully considered.

What were the survey sample designs? Unlike the previous considerations, consideration of sample designs does not lead to a clear yes or no decision regarding the combining of surveys. Sample design considerations provide information that helps to inform and structure the combining of survey data. Presumably, researchers will first consider issues related to the sample frame, survey data, and survey administration time differences before considering the sample design, and thus have some indication about the possibility of combining data from multiple surveys. In reviewing survey sample designs it is desirable to know the mechanics of sample selection (e.g., a random digit dialing (RDD) telephone survey, a multistage sample beginning with selection of a geographic area then selection of housing units then selection of respondents, or a sample selected from a preexisting list). This knowledge along with the rates of selection used for each survey provides an indication regarding the similarity of sample designs. Generally, the more similar the sample designs, the more defensible the decision to combine. This is not to say that surveys with very different sample designs cannot be combined, only that the case for combining may be more difficult to justify.

Combining Survey Data Sets: Technical Issues

This section addresses the mechanics of constructing weights when combining multiple weighted surveys that cover a common population. Rather than enumerating the range of possible combining strategies, this section presents a general strategy of scaling the weights that optimizes the precision of combined survey estimates overall. This strategy is meant as an example of statistical manipulation that will be required when weighted data sets are combined. It is <u>not</u> meant to serve as a guide for use with every set of data sources. Additional statistical consulting should be sought before data are combined and analysis begins.

Scaling Weights. It is possible, in some configurations of sample and weighting designs, that estimates for combined data sets may have less precision than those for each of the surveys being combined.[10] This condition is clearly at odds with the common reason for combining surveys – to increase the statistical power and precision of combined survey estimates over those for the individual surveys. This may occur because different sample designs and survey weighting strategies can affect the precision of combined survey estimates. Consider the circumstance of combining two surveys of a common population. Each survey was weighted to reflect the survey population. If the data are simply combined with no adjustment to weights, the combined weighted data set would yield population counts twice that of the population. Furthermore, this combined data set would not take into account response rates or survey design effects resulting from the sampling strategies. The result of combining these data sets could lead to less statistical power and less precise survey estimates rather than more power and more precise survey estimates.

Combining survey data sets covering a common population requires an adjustment or scaling of the survey weights to reflect the population of interest as well as the precision of the individual surveys. In general terms, for two surveys, the scaled weights for common domains are:

$$w_{ci} = \begin{cases} \alpha w_{1i} \text{ if } w_{1i} \text{ is the weight from Survey 1, and} \\ ((1-\alpha) w_{2i} \text{ if } w_{2i} \text{ is the weight from Survey 2} \end{cases}$$

where w_{ci} is the combined weight and α is the scaling factor subject to the restriction $0 \leq \alpha \geq 1$. As a result of the boundaries for α, $(\alpha + (1-\alpha)) = 1$, so the combined weighted estimate will equal the population total. Where the domains of the two surveys are not common, for example, the weight for each survey can be used without scaling.

There are several ways to select values for α. Among the most common are setting α as a function of each survey's sample size or setting α so that combined survey estimates have the smallest possible variance. The method recommended here minimizes the variance of combined survey estimates. In this method, α is set as a function of the variability of weights for the two surveys. Basically, the survey with the greatest precision is given greater relative influence in the scaling of weights. The formula used for calculating α is:

[10] See, for example, Cervantes, Jones, and Wilson (2005), *2005 Workplace and Equal Opportunity Survey of Active Duty Members: Statistical Methodology Report.*

$$\alpha = 1 - \frac{DEFF(w_{1i})}{DEFF(w_{1i}) + DEFF(w_{2i})}$$

where $DEFF(w_{1i})$ and $DEFF(w_{2i})$ are the respective design effects due to unequal weighting for Survey 1 and Survey 2 (it makes no difference which *DEFF* is used in the numerator). Calculation of the design effect due to unequal weighting for a survey is accomplished using the following formula:

$$DEFF(w_{1i}) = 1 + (CV(w_{1i})/100)^2$$

where $CV(w_{1i})$ is the coefficient of variation (CV) for the weights in Survey 1. The CV is calculated by dividing the standard deviation of the weights by the mean of the weights. This measure of relative dispersion is usually expressed as a percentage so the standard deviation divided by the mean is usually multiplied by 100.

Summary

The determination of whether to combine surveys for the construction of a more robust combined database is initially a substantive, not statistical, question. Answers to these questions, however, require considerable content familiarity. The initial questions presented in this appendix regard nuances of sample frame construction, questionnaire characteristics (question wording, response categories, and question placement context), dates of survey administration, and sample design. These issues are best evaluated by persons familiar with the research questions at hand and the population(s) of interest. The second section of the appendix presents an example of an approach for adjusting the combined weight when two weighted surveys covering the same population are combined. This example provides information for a general approach that will ensure that combination results in increased precision of the survey estimates and, more importantly, demonstrates the level of statistical expertise that may be required to correctly aggregate two data weighted data sources.

Appendix C: Review Protocol

Reviewer: Name of Data Source being Reviewed:	
Sponsoring Agency:	
Policy issue relevant to data source *(reference to policy document)*:	
Brief description of data set: *(Include brief statement of purpose, common uses, and all major topics covered in the data set.)*	
1. Determine if data source should be included in catalog.	
A. Does the data source allow you to identify people who are AI/AN/NA? □ Yes □ No (See below) If not, can you focus on a specific geographic region with a large AI/AN/NA population? □ Yes □ No (Exclusion Criteria)	
B. Is a data set available to researchers? □ Yes, available to public □ Yes, available through data use agreement (please explain process): □ Data are only available for government employees and contractors □ Aggregate program report is available on paper only (Exclusion Criteria) □ Data set is not available, but can request analyses □ Data set is not available, but published tables and reports are available □ Data set is not available at all (Exclusion Criteria) □ Other, explain:	Is there a cost associated with use of the data set? □ Yes (Cost: $_____no cost) □ No

Reviewer: Name of Data Source being Reviewed:
C. Please list the relevant contact and location information regarding this data source. Please include contact information for any required data use agreements and the location of the actual data.
D. **If data collection was privately funded**, were the data collected primarily to fill advocacy needs, lobbying purposes, or to support corporate interests? □ Yes (Exclusion Criteria), explain below: □ No □ Data are publicly funded
E. Total unweighted AI/AN/NA population in data source *(please include all possible breakdowns)*: **If unweighted n < 100**, would it be possible to combine this data source across multiple iterations of data collection to increase the size of the population? □ Yes , explain below: □ No (Exclusion Criteria)
F. What is the coverage of the data set? *(Check only one)* □ National □ Smaller geographic area of clear relevance to AI/AN/NA population □ A specific AI/AN/NA subpopulation only (e.g., single tribe, Pacific Islanders) □ None of the above (Exclusion Criteria)
G. What is the timeframe of the data set? □ Mid-1990s or later (can be series that began earlier as long as continues past mid-1990s) □ Earlier than mid-1990s If earlier than mid-1990s, can a strong argument be made to include this data set? □ Yes Explain: □ No (Exclusion Criteria)

Reviewer: Name of Data Source being Reviewed:	
H. Is this a survey? □ Yes □ No	For survey data, is detailed documentation available (either hard-copy or electronically) for: Instruments used for data collection (examples of question phrasing)? □ Yes □ No (Exclusion Criteria) Data limitations (e.g., coverage, data completeness, response rates)? □ Yes □ No (Exclusion Criteria) Sampling design techniques? □ Yes □ No (Exclusion Criteria) □ Not Applicable (e.g., Census) Weighting methods? □ Yes □ No (Exclusion Criteria) □ Not Applicable (data were not weighted) Contact information for program officials? □ Yes □ No (Exclusion Criteria)
I. Is this program reporting data? *(Include program enrollment)* □ Yes □ No	For program-level data, is detailed documentation available (either hard-copy or electronically) for: Contact information for program officials (e.g., primary and secondary contacts, telephone numbers and email addresses)? □ Yes □ No (Exclusion Criteria) Data limitations (e.g., coverage, data completeness)? □ Yes □ No
J. Is this a registry (e.g., vital statistics, state cancer registry, etc.)? □ Yes □ No	
K. Is this a qualitative study (e.g., case studies, focus groups, cognitive interviews)? □ Yes (Exclusion Criteria) □ No	

STOP AND DETERMINE WHETHER THE DATA SET WILL BE INCLUDED IN THE CATALOG BEFORE CONTINUING

Based on initial review, is this data source being profiled for possible inclusion in the catalog? □ Yes □ No, explain:
If not in the data catalog, do you recommend that this data source be included in the list of unreviewed data sources? *(Possible exclusions include qualitative, experimental, or methodological studies or studies with a significant lack of documentation.)* □ Yes, explain: □ No, explain:

Reviewer: Name of Data Source being Reviewed:
2. **Describe data source to be included in the catalog.**
A. How are AI/AN/NA individuals identified in the data set? *(Please note if separate multi-racial categories exist (such as American Indian and other vs. American Indian alone.) Also, if instructions for self-reporting are available, please include them here.)*
B. Does the data source allow you to identify members of a <u>federally recognized</u> tribe? □ Yes (Please explain how they are distinguished:) □ No Does the data source allow you to identify members of a <u>state-recognized</u> tribe? □ Yes (Please explain how they are distinguished:) □ No
C. What geographic area(s) does the study cover? □ National □ State(s), specify: □ County(ies), specify: □ Reservation(s), specify: □ Other region(s), specify:
D. What is the unit of analysis? □ Individual □ Geographical area (please specify:) □ Household or family unit (please specify definition of household:) □ Other, specify:
E. How are geographic areas identified? *(e.g., zip codes, census tracts, counties, county subdivisions, reservations, frontier areas)*

Reviewer: Name of Data Source being Reviewed:
F. What level of geographic analysis is possible?
G. What AI/AN/NA subpopulations are identified? □ American Indian alone □ Alaska Native alone □ Native Hawaiian alone □ Pacific Islander alone □ Specific tribal affiliation: _____
H. Is participation in this data collection mandatory or optional? □ Mandatory □ Optional, with incentives □ Optional, without incentives
I. Using the checklist below, please indicate all policy issues that are covered in this data source: **Health Status and Health Policy Issues** □ Measurement of health status (e.g., self-reported, % with disability, mortality/morbidity rates, trends over time) □ Disease-specific measurements (e.g., % with diabetes, TB, STDs, cancer, etc.) □ Key health disparities of priority interest (e.g., prenatal care/birth outcomes, cancer mortality, substance abuse/alcohol use/mental health/suicide) □ Factors contributing to measured health disparities (e.g., access to health care, insurance coverage, utilization rates, health care financing, socioeconomic factors, preventative measures (i.e., immunization rates) □ Identification of evidence-based practices and programs that address causes of health disparities, result in positive health outcomes, and are generalizable/replicable □ Role of traditional medicine in AI/AN/NA communities **Economic Well-being** □ Income status (e.g., household income/poverty status, per capita income) □ Unemployment rates □ Economic assistance program participation rates (e.g., TANF/TTANF, food stamps) □ Economic opportunity (e.g., number of businesses/jobs, work history) □ Measurement of economic/employment disparities between AI/AN/NA and general population □ Factors contributing to economic disparities (e.g. lack of child care arrangement, transportation barriers) □ Identification of evidence-based practices and programs for reducing economic disparities and are generalizable/replicable

Reviewer:
Name of Data Source being Reviewed:

(Policy Issues Checklist continued)

Education Levels and Opportunities

□ Educational attainment (e.g., last grade completed, literacy/numeracy skills)
□ Educational opportunities (e.g., Head Start, special education programs, school financing)
□ Factors contributing to educational disparities (e.g., parents' education level, average education in city/county, education spending per capita, and other socioeconomic factors)
□ Identification of evidence-based practices and programs that produce positive educational outcomes and are generalizable/replicable

Family Well-being

□ Measures of well-being for families/households (e.g., families with low income levels, homeless families, teen pregnancy/birthrates, household size and composition)
□ Factors contributing to well-being disparities of families (e.g., socioeconomic factors, education levels of family adults, housing quality, public transportation availability)
□ Identification of evidence-based practices and programs that improve family well-being and are generalizable/replicable

Child Well-being

□ Measures of well-being for children (e.g., children in foster care, incarcerated children)
□ Factors contributing to well-being disparities of children (household composition, martial status of parents, foster care placement)
□ Identification of evidence-based practices and programs that improve child well-being and are generalizable/replicable

Elder Well-being

□ Measures of well-being for elders (e.g., elders with low income levels, homeless elders, elder abuse)
□ Factors contributing to well-being disparities of elders (e.g., socioeconomic factors, living arrangements, ADL/IADL, family members in proximity, services available/used (Meals on Wheels/elder transportation)
□ Identification of evidence-based practices and programs that improve elder well-being and are generalizable/replicable

Housing Issues

□ Housing quality (e.g., rooms per person, running water, electricity, heat, age of building)
□ Type of housing
□ Housing ownership
□ Rental unit quality and cost
□ Homelessness

Demographic and Economic Indicators (e.g., age distribution, marital status, household composition)

Reviewer:
Name of Data Source being Reviewed:

(Policy Issues Checklist continued)

Transportation Quality and Availability Issues

□ Transportation quality (e.g., maintenance of roads, availability of paved roads, child restraint laws)
□ Transportation availability (e.g., availability of public transit, vehicle ownership per household)

Justice System Issues

□ Rates of involvement with justice system (e.g., arrest/conviction/probation/parole rates)
□ Differences in resolution of arrest, by type of court system (e.g., federal, tribal, state, local)
□ Lifetime probability of being a victim of a violent crime
□ Lifetime probability of being a victim of a non-violent crime
□ Domestic violence rates
□ Child maltreatment rates
□ Factors contributing to disparities in involvement with justice system and outcomes (e.g., family stability/foster care placement, family members' history of legal system involvement, race/ethnicity, truancy history)
□ Identification of evidence-based practices or programs to reduce involvement with justice system, reduce recidivism, and are generalizable/replicable

Military Service/Veterans' Issues

□ Military service rates (e.g., % served in military, % retired from military with benefits)
□ Eligibility and use of veterans administration health facilities
□ Eligibility and use of other veterans administration benefits (e.g. housing loans, educational benefits)

J. Describe data collection methodology (e.g., mode).

K. What is the response rate for this study? *(If possible, please provide any details on how the response rate was calculated).*

L. How frequently are the data collected?

□ One-time effort
Year of data collection:

□ More than once *(e.g., annual study, semi-annual study, longitudinal study, irregular points of data collection)*
Please describe *(include the years the data are available for and any plans for future data collections)*:

Reviewer: Name of Data Source being Reviewed:	
SURVEYS ONLY	M. Describe sampling method, if any. (e.g., simple random sample, two-stage stratified sample) N. Was the AI/AN/NA population oversampled? □ No □ Yes (Please describe:) O. Does the methodology report include any of the following (please provide as much information as possible): □ Effective sample size (n=_____) □ Design effects (deff): _____ □ Standard errors for estimates: _____ □ Description of available power to detect differences *(This may be found in the sample section of a methodology report. Please specify variables used for calculation and for which groups differences are to be detected)*
P. If multiple years of data are available and are suitable for aggregation, what would be the result of the aggregation (e.g., sample size)?	
Q. Summarize the strengths of the data source. *(Include areas such as coverage in terms of sample size and content about AI/AN/NA population, relevance to policy areas, amount of missing data, response rate, weighting, quality of documentation, psychometric assessments)* <u>Key Issues</u> □ Large number of AI/AN/NA respondents □ Relevance to key policy issues, including: □ Health □ Child welfare □ Multiple years of data are available <u>Additional Strengths:</u>	

Reviewer: Name of Data Source being Reviewed:
R. Summarize the limitations of the data source. *(Include areas such as coverage of AI/AN/NA population, content pertaining to the target population, relevance to policy areas, amount of missing data, response rate, weighting, quality of documentation)* <u>Key Issues</u> □ Small number of AI/AN/NA respondents □ Lack of relevance to key policy issues □ Multiple years of data are <u>not</u> available <u>Additional Limitations:</u>
S. Was this data collection effort authorized by any specific legislation? *(Answer yes if Congress created a mandate to authorize this specific data collection effort.)* □ Yes (Explain:) □ No
T. Other comments?
U. Document materials in project files related to this data source, or related web links.

Appendix D: List of Sponsoring Agencies

Sponsor	Data Set
National Science Foundation	• Panel Study of Income Dynamics (PSID)
State of Hawaii Department of Health (DOH)	• Hawaii Health Survey (HHS)
State of Washington	• Washington State Population Survey (WSPS)
The Urban Institute	• National Survey of America's Families (NSAF)
U.S. Department of Agriculture (USDA)	• Census of Agriculture (2002) • Food Stamp Quality Control Database (FSPQC)
U.S. Department of Commerce	• American Community Survey (ACS) • Census 2000 • Census 2000 – American Indian and Alaska Native Summary File • Current Population Survey (CPS)* • National Longitudinal Mortality Study (NLMS)* • Small Area Income and Poverty Estimates (SAIPE) • Survey of Program Dynamics
U.S. Department of Education (DoE)	• Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) • Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K) • Integrated Postsecondary Education Data System (IPEDS) • National Assessment of Adult Literacy (NAAL) • National Household Education Surveys Program (NHES) • National Indian Education Study (NIES)
U.S. Department of Justice (DoJ)	• Annual Survey of Jails (ASJ) • Census of Jails • Census of Tribal Justice Agencies in Indian Country (2002) • National Crime Victimization Survey (NCVS) • Survey of Jails in Indian Country (SJIC) • Uniform Crime Reports • Youth Gangs in Indian Country
U.S. Department of Labor (DoL)	• Consumer Expenditure Surveys (CE) Interview and Diary Surveys • Current Population Survey (CPS)*
U.S. Department of Transportation	• National Household Travel Survey (NHTS)
U.S. Department of Veterans Affairs	• National Survey of Veterans (NSV)
UCLA Center for Health Policy Research	• California Health Interview Survey (CHIS)
U.S. Department of Housing and Urban Development (HUD)	• A Picture of Subsidized Households (1998) • American Housing Survey (2003) • American Housing Survey: Metropolitan Surveys

Sponsor	Data Set
U.S. Department of Health and Human Services (DHHS)	• Adoption and Foster Care Analysis and Reporting System (AFCARS) • Behavioral Risk Factor Surveillance System (BRFSS) • CAHPS Health Plan Survey Response Data • Head Start Program Information Report • Health Behavior in School-aged Children (HBSC) • Health Information National Trends Survey (HINTS) • Health and Retirement Study (HRS) • Medicaid Analytic Extract (MAX) • Medical Expenditure Panel Survey (MEPS) • Medicare Denominator File • Medicare Utilization – Standard Analytic Files (SAFs) • National Aging Program Information Systems (NAPIS) State Performance Reports • National Ambulatory Medical Care Survey (NAMCS) • National Child Abuse and Neglect Data System (NCANDS) • National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) • National Health Interview Survey (NHIS) • National Hospital Ambulatory Medical Care Survey (NHAMCS) • National Longitudinal Mortality Study (NLMS)* • National Mortality Followback Survey (NMFS) • National Survey of Child and Adolescent Well-being (NSCAW) • National Survey on Drug Use and Health (NSDUH) • National Survey of Family Growth (NSFG) • National Vital Statistics System: Linked Birth-Infant Death (NVSS-I) • National Vital Statistics System: Mortality (NVSS-M) • National Vital Statistics System: Natality (NVSS-N) • Pediatric Nutrition Surveillance System (PedNSS) • Pregnancy Nutrition Surveillance System (PNSS) • Pregnancy Risk Assessment Monitoring System (PRAMS) • Resource and Patient Management System (RPMS) and National Patient Information Reporting System (NPIRS) • Runaway and Homeless Youth Management Information System (RHYMIS) • Surveillance, Epidemiology, and End Results (SEER) • Temporary Assistance for Needy Families (TANF) and Tribal TANF • Tobacco Use Supplement to the Current Population Survey (TUS-CPS) • Treatment Episode Data Set (TEDS) • United States Renal Data System (USRDS) • Youth Risk Behavior Surveillance System (YRBSS)

* Denotes data sets that have more than one sponsor.