

GAPS AND STRATEGIES FOR IMPROVING AI/AN/NA DATA

FINAL REPORT

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

This report is a component of ASPE's Data on Health and Well-being of American Indians, Alaska Natives, and Other Native Americans project.

Prepared by:

Westat
1650 Research Boulevard
Rockville, MD 20850

January 2007

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/07/AI-AN-NA-data-gaps>

ACKNOWLEDGMENTS

This report was prepared by Westat and the American Indian Health Research Program, Black Hills State University. The principal authors are Kathryn Langwell, Cynthia Helba, and Craig Love. 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 for the project. Canta Pian, Division Director, Economic Support for Families/ASPE, also 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 report. Additionally, several members of an American Indian, Alaska Native, and other Native American workgroup provided input and review during the course of the project including David Wong representing the IHS Division of Epidemiology and Disease Prevention, 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.

Numerous knowledgeable individuals at DHHS, the Census Bureau, 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, databases and initiatives that are underway and planned that will improve AI/AN/NA data. 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.

TABLE OF CONTENTS

	<u>Page</u>
EXECUTIVE SUMMARY	1
1 INTRODUCTION	8
2 BACKGROUND AND REVIEW OF PRIOR STUDIES.....	9
2.1 Overview of Study Populations	9
2.1.1 American Indians and Alaska Natives.....	9
2.1.2 Native Hawaiians and Other Pacific Islanders.....	11
2.2 Findings of Prior Studies and Reports	13
2.2.1 Availability and Quality of Data on AI/AN/NA Health and Well-being.....	13
2.2.2 Factors Contributing to Gaps in AI/AN/NA Data	15
2.2.3 Strategies for Improving AI/AN/NA Data Availability and Quality..	16
3 METHODS	18
3.1 Review of the Literature and Reports	19
3.2 Analysis of Data Catalog Information	19
3.3 Key Informant Interviews	20
3.4 Limitations	21
4 SUMMARY OF DATA AVAILABILITY AND GAPS IN SELECTED DATA SETS, BY POLICY AREA AND POPULATION GROUP.....	21
4.1 Data Availability, by Policy Area and Population Group.....	22
4.2 Sample Sizes 200 or Greater, by Policy Area and Population Group	24
4.3 Geographic Information Available, by Primary Policy Area and Population Group	25
4.4 Data Gaps by Policy Area and Population Group	26
4.5 Discussion	28
5 STRATEGIES FOR IMPROVING DATA ON AI/AN HEALTH AND WELL-BEING	29
5.1 Issues that Contribute to Gaps in AI/AN/NA Data Availability and Quality.....	29
5.2 Strategies for Improving Data on AI/AN/NA Health and Well-being	31
5.3 Current and Planned Initiatives of DHHS and Other Federal Agencies that Would Improve Availability and Quality of Data on AI/AN/NA Health and Well-being	38
6 SUMMARY AND DISCUSSION.....	43
6.1 Summary of Findings.....	43
6.2 Conclusion	46

TABLE OF CONTENTS (continued)

List of Appendices

	<u>Page</u>
Appendix A – List of Policy Areas and Issues	A-1
Appendix B – Sample Sizes for Data Sets by Policy Area, for Each Population Group	B-1
Appendix C – List of Reports Reviewed	C-1
Appendix D – Key Informant Interview Guide	D-1

List of Tables

Table ES-1. Data Gaps, by Policy Area and Population Group–Data Sets With 200 or More Sample Size for Each Population Group	2
Table 1. Number of Data Sets that Identify Specific Population Groups, by Policy Area.....	22
Table 2. Number of Data Sets that Identify Specific Subpopulation Groups	23
Table 3. Number of Data Sets with Sample Size 200 or More by Population Group	24
Table 4. Data Gaps, by Policy Area and Population Group–All Data Sets	27
Table 5. Data Gaps, by Policy Area and Population Group–Data Sets With 200 or More Sample Size for Each Population Group	28
Table 6. Summary of Strategies, Potential Costs, and Current/Planned Initiatives.....	47
Table B-1. Sample Sizes for the AI/AN Race Category.....	B-2
Table B-2. Sample Sizes for the AI Alone Race Category	B-5
Table B-3. Sample Sizes for the AN Alone Race Category	B-5
Table B-4. Sample Sizes for the NH/PI Race Category	B-5
Table B-5. Sample Sizes for the NH Alone Race Category	B-7
Table B-6. Sample Sizes for the PI Alone Race Category	B-8

EXECUTIVE SUMMARY

Background and Methods

This report is based on the findings from a study funded in 2005 by the Office of the Assistant Secretary for Planning and Evaluation (ASPE), U. S. Department of Health and Human Services (DHHS), entitled *Data on Health and Well-being of American Indians, Alaska Natives, and other Native Americans (AI/AN/NA)*. The objectives of this study were to:

- Systematically review federal and other data sets to assess the extent to which data on health and well-being of the American Indian/Alaska Native (AI/AN) and Native Hawaiian/Pacific Islander (NH/PI) populations are available and adequate for identifying and monitoring measures of health and well-being for these groups;
- Prepare a data catalog with information on available data sets with adequate samples of the AI/AN and NH/PI populations for identifying and monitoring health and well-being and for supporting research that could provide useful insights into factors that contribute to disparities; and
- Identify gaps in the available data, current initiatives underway within DHHS and other federal agencies that may reduce these gaps, and strategies that, if implemented, could improve the availability and quality of data on AI/AN/NA health and well-being.

To address the third objective, this report identifies gaps in AI/AN/NA data on health and well-being, strategies for improving AI/AN/NA data availability and quality, and some current initiatives underway within DHHS and other federal agencies that are intended to improve these data.

The findings in this report are drawn from a review of the literature and available reports on AI/AN/NA data, analysis of detailed profiles of 67 data sources that were included in the *Data Catalog on AI/AN/NA Health and Well-being*, an earlier product of this study, and interviews with 13 knowledgeable individuals who provided information on current initiatives underway and additional strategies that could improve the availability and quality of data on AI/AN/NA health and well-being. In developing the data catalog, we identified a set of policy areas relevant to AI/AN/NA health¹ and well-being, including child well-being, economic well-being, educational achievement and opportunities, elder well-being, family well-being, housing indicators, justice system indicators, military/veterans' issues, and transportation quality and availability. The data catalog analysis focuses on the availability of data sets to address each of the identified policy areas.

¹ Health policy topics considered in the catalog include such areas as health status (e.g., self-reported health status, disability rates, mortality/morbidity rates, trends over time); disease-specific prevalence and incidence (e.g., prevalence of diabetes, tuberculosis, sexually transmitted diseases); 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, utilization rates, health insurance coverage, health care financing, socioeconomic factors, preventative measures); evidence-based practices and programs that address causes of health disparities, result in positive health outcomes, and are generalizable/replicable; and the role of traditional medicine in AI/AN/NA communities.

The methodology used for this study has some limitations, most notably, only 67 data sets were thoroughly reviewed; it was not possible to fully examine the quality of the data sets; and only a limited number of knowledgeable individuals were interviewed to obtain information on current and planned initiatives to improve AI/AN/NA data. Although only 67 data sets comprise the foundation for this report, it is unlikely that there are many other federal or other publicly available databases that have sufficient sample sizes of AI/AN/NA and include measures of health and well-being that would substantially affect the findings of this analysis. (There are of course still yet fewer data sets containing information on individual tribes or other subgroups.) Because it was not possible for us to fully examine the quality of individual databases, it is possible that analysts and researchers who use these data sets may identify quality issues that will result in additional gaps pertaining to the health and well-being policy areas. Finally, it is very likely that the limited number of knowledgeable people who were interviewed were not aware of all of the current and planned initiatives within DHHS and other organizations. Thus, there likely are more activities underway and planned than are presented in this report that could contribute to improved data on AI/AN/NA health and well-being. Given these caveats, this report provides the reader an overview of gaps in AI/AN/NA data, strategies to improve data, and some information on current initiatives that may be valuable in leading to improved data in the future.

Findings: Data Availability and Gaps, By Policy Area and Population Groups

Our analysis of the 67 data catalog profiles indicates that there are more than two data sets available for each of the major policy areas listed earlier, both for the combined AI/AN and combined NH/PI population groups. There are fewer data sets available that permit examination separately of the AI, AN, NH, and PI population groups. When the available data sets are limited to those with sample sizes of 200 or greater, the number of data sets available to examine primary policy issues is reduced (Table ES-1). For most policy issues and population groups, however, there are data sets that can be used to produce information without requiring aggregation of data or special statistical techniques.

**Table ES-1. Data Gaps, by Policy Area and Population Group—
Data Sets With 200 or More Sample Size for Each Population Group**

Policy Area	AI/AN	AI	AN	NH/PI	NH	PI
Child Well-being	>2	Gap	Gap	>2	Gap	Gap
Economic Well-being	>2	Limited	Limited	>2	Limited	Limited
Education	>2	Limited	Limited	Limited	Limited	Limited
Elder Well-being	>2	Gap	Gap	Limited	Gap	Gap
Family Well-being	>2	Limited	Limited	>2	Limited	Limited
Health	>2	>2	>2	>2	>2	Limited
Housing	>2	Limited	Limited	>2	Limited	Limited
Justice System	>2	Gap	Gap	Limited	Limited	Gap
Military/Veterans	Limited	Gap	Gap	Gap	Gap	Gap
Transportation	>2	Limited	Limited	Limited	Limited	Limited

Note: “>2” indicates that there are more than two data sets that permit analysis of the policy area for the specific population group. “Gap” indicates that no data set was identified for that policy area for the specific population group. “Limited” indicates that only one or two data sets were identified for the policy area for the specific population group.

Major findings of this analysis of data sets with at least 200 sample size include:

- For the combined AI/AN population, there are more than two data sets for each policy area, with the exception of military/veterans issues.
- For the combined NH/PI population, five of the 10 policy areas are covered by data sets; there are limited data (i.e., two or fewer data sets) for education, elder well-being, justice system issues, and transportation issues; and no data sets are available for examination of military/veterans issues.
- For the AI and AN populations separately, there are more than two data sets available to examine health issues, no data sets available to examine child well-being, elder well-being, justice system issues, and military/veterans issues and limited data available for each of the other policy areas.
- For the NH group separately, there are more than two data sets available to examine health issues, limited data available to examine economic well-being, education, family well-being, housing, justice system issues, and transportation issues. No data sets are available to examine child-well-being, elder well-being, or military/veterans issues.
- For the PI group separately, there are gaps in data availability for child-well-being, elder well-being, justice system issues, and military/veterans issues and limited data available for examination of each of the six other policy areas.

While we found no gaps in the available data for the combined AI/AN group and only one gap for the combined NH/PI data, there are a substantial number of policy areas for which data are not available to examine health and well-being for the separate AI, AN, NH, and PI groups. It is important, also, to remember that the data sets examined may vary in the indicators available to examine specific issues within a policy area. For some policy areas and issues, there may be inadequate data for analysis of disparities in health and well-being for these population groups, by demographic, geographic, or economic characteristics, even though some data on the issue/policy area may be collected and usable for aggregate analysis.

Findings: Strategies and Current/Planned Initiatives for Improving AI/AN/NA Data Availability

Information on issues that affect AI/AN/NA data availability and quality and strategies to address these issues has been drawn from the literature reviewed, from the discussions held with DHHS staff and other knowledgeable individuals, and from our own assessment of the information obtained and reviewed for this study. We also obtained information from the discussions on current and planned initiatives within DHHS and other federal agencies that may, over the next few years, result in improved data and greater data availability to address a range of health and well-being issues for the AI/AN/NA population groups. Below, the issues that affect data availability and quality and the strategies for improving data that were identified are

summarized. Following that we briefly describe current and planned initiatives within DHHS and other federal agencies related to these strategies.

Issue 1: Small Population Size: AI/AN constitute 1.5 percent² of the U.S. population and NH/PI constitute about 0.3 percent of the U.S. population. Because of their small numbers, many federal and other surveys do not have sufficient sample sizes to obtain representative data for these groups. Several strategies were identified that would address the small population issue, including: 1.a) increase sample sizes of federal surveys to ensure that sufficient samples are obtained to permit analysis of AI/AN and NH/PI health and well-being issues; 1.b) over-sample the AI/AN and NH/PI population in federal surveys in order to obtain sufficient sample sizes for these groups; 1.c) aggregate multiple years of data from surveys that are conducted annually to obtain sufficient AI/AN and NH/PI sample sizes; 1.d) fund and conduct additional methodological studies to develop new approaches to using small samples for estimating AI/AN and NH/PI health and well-being; and 1.e) explore the potential for developing partnerships with AI/AN and NH/PI communities to conduct local area versions of major federal surveys that could supplement national survey data.

Issue 2: Geographic Dispersion and Rural Concentration of the AI/AN and NH/PI Populations: The AI/AN and NH/PI populations are geographically concentrated in rural areas and are geographically dispersed among the general population when they are not residing on reservations or in Native Hawaiian homelands. Rural concentration increases the barriers to inclusion in nationally representative survey samples, and geographic dispersion makes it more difficult to develop sampling strategies that would increase AI/AN/NA representation in national surveys. Strategies that were identified to address these issues include: 2.a) develop and implement new methods for cost-effective identification of AI/AN and NH/PI who reside outside reservations/homelands to permit cost-effective sampling; 2.b) increase rural sample sizes or over-sample AI/AN/NA in these areas to ensure that data on AI/AN residing on reservations and NH residing on homelands are available; and 2.c) fund and conduct research to develop new methodological approaches that would make maximum use of existing data from rural areas, while protecting privacy and confidentiality of respondents.

Issue 3: Misclassification of Race: Race may be reported by officials other than the individual or family members, particularly in vital statistics and administrative databases.³ Studies have shown that misreporting of race may occur often and that the magnitude of misreporting varies geographically. One strategy was identified that would address this issue: 3.a) develop new requirements and guidelines for accurate reporting of race/ethnicity on vital statistics records and administrative databases and provide training for those who are responsible for reporting.

Issue 4: Lack of/Inconsistent Collection of Race Identifiers in Some Data Sources: Race may not be identified or reported on some state and local databases and different states may report race using different racial identifiers. One strategy was identified to address this issue: 4.a) develop guidelines and encourage or require state agencies to collect uniform race/ethnicity data for all federally funded and federal-state jointly funded programs.

² The 2000 Census reports that 0.9 percent of the population lists American Indian/Alaska Native as their only race, while 1.5 percent indicate AI/AN race either alone or in combination with one or more other races.

³ Here, administrative databases include program-specific databases, disease registries, and disease surveillance systems.

Issue 5: Inadequate Racial Representation, Limited Response Rates, and Question Interpretation:

The sampling strategies used for some surveys (e.g., emphasis on urban samples rather than rural) may reduce the likelihood that AI/AN and NH/PI will be representatively included in the survey sample frame. In addition, the use of certain survey methods (e.g., by telephone) may reduce the number of AI/AN and NH/PI due to an inappropriate fit of the method with the circumstances of the target population. Finally, there is also some evidence that cultural and language differences of these populations may result in interpretation and response to survey questions that differ from those of the general population. Strategies that have been identified to address these issues include: 5.a) review sample designs for existing surveys to identify their potential to include AI/AN and NH/PI representatively in surveys and develop approaches that could increase representation; 5.b) fund and conduct additional research on approaches that could increase AI/AN and NH/PI response rates and test/implement these approaches; and 5.c) conduct ongoing cognitive testing of current and new survey instruments and questions to assess the influence of cultural differences in question interpretation and develop alternative wording, if necessary.

Issue 6: Exclusion of Pacific Insular Areas from Some Surveys: The Pacific Insular Areas include dependent territories of the U.S. (Guam, American Samoa, and the Northern Mariana Islands) and the Freely Associated States for which the U.S. provides defense, funding grants, and social services to its citizens (Micronesia, Palau, and the Marshall Islands). Because these areas are very distant from the U.S. mainland and are scattered across a very large geographic area in the Pacific Ocean, most national surveys do not include these territories and Freely Associated States. One strategy was identified to remedy this issue: 6.a) include Pacific Insular Areas in most relevant surveys and other data collection, either routinely or periodically.

Issue 7: Inadequate Collection of Data on AI/AN/NA Subgroups: There are few data sets available to provide information on AI, AN, NH, and PI as separate groups, although there is evidence that there may be substantial differences in health and well-being among these groups. Similarly, there is little data available on members of federally recognized tribes, despite the unique government-to-government relationship and federal obligations established by treaties with these tribes. One strategy was identified to address this issue: 7.a) refine racial identifiers to better focus on subgroups of AI/AN and NH/PI populations and design and implement data collections to obtain sample sizes sufficient for analysis of health and well-being for specific subgroups.

Discussions with a limited number of knowledgeable individuals within DHHS, Census, and other organizations identified a number of current or planned initiatives that would improve AI/AN/NA data availability and quality. All but three of the strategies identified above had current or planned initiatives underway that would, at least partially, address the associated issue. The three strategies for which there were no current or planned initiatives underway are: 1) increasing sample sizes of relevant federal surveys; 2) over-sampling of AI/AN/NA within existing relevant federal surveys; and 3) increasing sample sizes for AI/AN/NA residing on reservations, AN villages, and NH homelands. Each of these strategies would potentially require substantial additional costs to implement. In addition, the strategy to improve PI data by including the Pacific Insular Areas in relevant federal surveys is under study, but is not actually being implemented as a strategy. Again, the cost of implementation of this strategy is likely to be high, depending on whether these areas were included periodically or routinely in federal

surveys. Current or planned initiatives are underway for all of the other strategies, either within federal agencies, states, or through academic-native community partnerships. None of these strategies would likely involve high additional costs, and most could be implemented for relatively modest costs.

Summary and Conclusions

The review of the availability of data on AI/AN/NA health and well-being identified a number of findings, including:

- There are some data available on health and well-being of the combined AI/AN and NH/PI populations for all of the health and well-being policy areas that were the focus of this study.
- There are fewer data available, however, when the data sets examined were restricted to those that had sample sizes for each combined group of at least 200—a number that would permit analysis of the health and well-being issue for some characteristics (e.g., age, gender). Specifically, for the AI/AN combined group, there are limited data (i.e., only one or two data sets) to examine military/veterans issues. For the NH/PI combined group there are limited data to examine education, elder well-being, justice systems issues, and transportation issues, and there are no available data sets to examine military/veterans issues.
- Data to examine measures of health and well-being for the separate AI, AN, NH, and PI groups are much less available than for the AI/AN combined and NH/PI combined groups. Given the evidence that there may be substantial differences in health and well-being among these subgroups, the lack of available data to measure these differences is an important issue.

The study identified a number of issues that contribute to the availability of data on health and well-being of the AI/AN/NA population and strategies for addressing these issues to improve data availability and quality. Interviews with a small number of knowledgeable individuals were conducted to identify current initiatives underway within DHHS and other federal agencies that would potentially result in improvements in availability and quality of AI/AN/NA data. A substantial number of initiatives are underway or planned in the future, and these current initiatives address all but three of the strategies that were identified in this report. The three strategies that have not been addressed within DHHS are those that would likely involve the highest costs to implement (e.g., increasing sample sizes of federal surveys to obtain larger numbers of AI/AN/NA respondents).

It is important to recognize that these current initiatives would advance the objective of improving data availability and quality, but individually would not fully address this problem. For instance, strategies that focus on research to identify new statistical approaches for small area/population estimation may produce a set of methodologies that could improve the usefulness of the available data, but implementation of those methodologies might involve complex and potentially costly software development that might be prohibitive. Similarly, research and cognitive testing to identify differences among populations in understanding and

interpretation of survey questions is a relatively low-cost strategy; however, implementation of strategies to address these differences and improve data quality may involve high additional costs.

A comprehensive approach to addressing the data gaps identified could involve the following steps:

- A DHHS-wide coordinated approach to implement many of the identified strategies across DHHS agencies that both use and collect data on health and well-being issues. Such a coordinated approach would involve soliciting information from these agencies on the key issues for measuring, monitoring, and conducting research on health and well-being disparities, sharing of information gained from current and planned initiatives across DHHS agencies, and identifying a process for determining specific approaches that would be adopted and used consistently in DHHS surveys and administrative databases. This could be accomplished through an incremental approach that begins with coordination among several key DHHS agencies, with the DHHS Data Council perhaps taking a lead role in the process. This effort could then be expanded over time to involve additional DHHS agencies in the process.
- Coordination and sharing of results of current initiatives among all federal agencies that use and collect data on AI/AN/NA health and well-being. This coordination and sharing of findings could lead to development of a process for adopting and implementing consistent data collection strategies to improve these data.
- Working with states to improve the completeness, accuracy, and consistency of collection of race identifiers to ensure that vital statistics and administrative databases contain accurate identification of AI/AN/NA people.
- Consultation and involvement of AI/AN/NA tribes and communities as partners in the process of improving these data, to ensure that strategies that are identified and implemented have the support of these populations and that the approaches that will be implemented also reflect the priorities of the population that will benefit from improved data.

If the coordination strategies outlined above are implemented, it is likely that the availability and quality of data on AI/AN/NA health and well-being will substantially improve. This outcome will require that a long-run focus be maintained on the need for and importance of data to measure, monitor, and analyze disparities in health and well-being of this population. In addition, consideration of the potential benefits relative to costs of alternative strategies may be useful for guiding the decisions that will support these important data improvements.

1 INTRODUCTION

The U.S. Department of Health and Human Services (DHHS) has a commitment to reduce health and well-being disparities between the majority population and racial and ethnic minorities. Recognizing the important role of data in identifying, monitoring, and reducing disparities, DHHS has initiated a series of activities over the past two decades to improve data collection concerning the health and well-being of racial and ethnic populations. These previous activities include work by 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.⁴ These earlier efforts identified the capacity to examine disparities for specific racial/ethnic minority groups within existing federal data sets, issues that contribute to lack of data and inadequate quality of data available, and strategies for improving these data.

While substantial progress has been made in addressing the issues identified as barriers to collecting and analyzing data for some racial/ethnic minority groups, the challenges posed by small population racial groups are particularly difficult to overcome. American Indians and Alaska Natives (AI/AN), Native Hawaiians (NH), and other Pacific Islanders (PI) as a group represent less than two percent of the U.S. population—AI/AN are 1.5 percent⁵ of the population and NH and PI combined are about 0.3 percent of the population. Data for these groups are less available than for other racial/ethnic groups, for a number of reasons. However, the limited data available suggest that AI/AN, NH, and PI experience substantial disparities that are greater, in some ways, than other racial/ethnic groups' experience.

This report is based on findings from a study funded in 2005 by the Office of the Assistant Secretary for Planning and Evaluation (ASPE), DHHS, entitled *Data on Health and Well-being of American Indians, Alaska Natives, and other Native Americans (AI/AN/NA)*. The objectives of this study were to:

- Systematically review federal and other data sets to assess the extent to which data on health and well-being of the AI/AN, NH, and PI populations are available and adequate for identifying and monitoring measures of health and well-being for these population groups;
- Prepare a data catalog with information on available data sets with adequate samples of the AI/AN and NH/PI populations for identifying and monitoring health and well-being and for research that could provide useful insights into factors that contribute to disparities; and
- Identify gaps in the available data, current initiatives underway within DHHS and other federal agencies that may reduce these gaps, and strategies that, if implemented,

⁴ *Improving the Collection and Use of Racial and Ethnic Data in HHS and Assessment of Major Federal Data Sets for Analyses of Hispanic and Asian Pacific Islander Subgroups and Native Americans* is available at <http://aspe.hhs.gov/datacncl/racerpt/> and <http://aspe.hhs.gov/HSP/minority-db00/task2/index.htm>.

⁵ The 2000 Census reports that 0.9 percent of the population lists American Indian/Alaska Native as their only race, while 1.5 percent indicate AI/AN race either alone or in combination with one or more other races.

could improve the availability and quality of data on AI/AN/NA health and well-being.

This report was written to fulfill the third objective stated above. It presents background information, a review of prior studies of AI/AN/NA data issues and a description of the methods used in this effort. Building on what was learned from compiling the data catalog, information is presented on health and well-being issues for which data are unavailable or inadequate for measuring, monitoring, and conducting research on health and well-being disparities of AI/AN, NH, and PI. In addition, current initiatives underway within DHHS and other federal agencies that are intended to improve data for these populations are identified and a number of potential strategies for further improving data are presented.

2 BACKGROUND AND REVIEW OF PRIOR STUDIES

This section provides a brief overview of the AI/AN and NH/PI populations, as context for the examination of health and well-being data availability and quality. This is followed by a review of prior studies and initiatives focused on assessing availability of data for these population groups and possible strategies for improving availability.

2.1 Overview of Study Populations

Much of the information in this section is drawn from 2000 Census reports on the AI/AN and NH/PI population. The 2000 Census implemented new racial/ethnic identification questions, based on Office of Management and Budget Directive 15 (OMB 15).⁶ OMB 15 required that the 2000 decennial census revise racial/ethnic questions to 1) permit respondents to self-identify more than one race; and 2) identify Native Hawaiians and Other Pacific Islanders as a new category, replacing the combined Asian/Native Hawaiian/Pacific Islander category.⁷ As a result of implementation of OMB 15 by the 2000 Census, NH/PI are separately identified and reported for the first time. However, the implementation of the requirement that respondents be permitted to self-identify multiple races resulted in a substantial increase in the number of people reporting AI/AN when both AI/AN alone or in combination with other races are considered. As a result, AI/AN Census data for 1990 and 2000 are not necessarily comparable.

2.1.1 American Indians and Alaska Natives

⁶ Office of Management and Budget, *Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity*. *Federal Register Notice*, October 30, 1997 (<http://www.whitehouse.gov/omb/fedreg/ombdir15.html>)

⁷ OMB 15 requires other federal programs to adopt these standards (and others in OMB 15 that are not directly related to the AI/AN and NH/PI population) no later than January 1, 2003. The directive covers federal household surveys, federal administrative forms and records, and other federal data collection. It does not, however, require state, local, or private organizations to implement these standards.

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.”⁸ 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. These population groups include:

- American Indians and Alaska Natives, combined or as separate groups, who are members of federally recognized tribes, regardless of residence.
- AI/AN, combined or as separate groups, who are members of federally recognized tribes and who reside on or near reservations or Alaska Native villages.
- AI/AN, combined or as separate groups, who are members of federally recognized tribes and who reside in urban areas or other locales that are not on or near reservations or Alaska Native villages.
- American Indians who are members of state-recognized tribes, who live on/near reservations or elsewhere.
- American Indians who are members of tribes that are not currently recognized by either the federal government or a state government.
- American Indians and Alaska Natives who self-identify as being of only AI/AN race. This definition includes members of federally recognized tribes, members of non-federally recognized tribes, and others who report AI/AN race but are not enrolled as members of a tribe.
- AI/AN who self-identify race as AI/AN, in combination with one or more other races. This definition includes members of federally recognized tribes, members of non-federally recognized tribes, and others who report AI/AN race but are not enrolled as members of a tribe.

American Indians and Alaska Natives who are members of federally recognized tribes, a subgroup of all who report AI/AN race on the U.S. Census, have a unique relationship with the federal government due to treaties that were signed by tribes and the federal government that established sovereignty of tribes and specified federal responsibilities to maintain the health and well-being of tribes.⁹ The federal government may provide health, housing, and other services to members of federally recognized tribes, although these services are generally provided to those tribal members who reside on or near reservations.

The Census Bureau reports that 2.5 million persons reported AI/AN race only on the 2000 Census, and over four million reported AI/AN race only or in combination with another race. Those reporting AI/AN race only are 0.9 percent of the population, while those reporting

⁸ U.S. Census Bureau, “The American Indian and Alaska Native Population: 2000.” Census 2000 Brief, February 2002, page 2.

⁹ Some tribes do not have treaties, but received federal recognition through other means such as executive orders or federal legislation.

AI/AN race only or in combination with another race are 1.5 percent of the U.S. population. Forty-three percent of the AI/AN population resides in the Western region of the U.S.; 31 percent in the South; 17 percent in the Midwest; and 9 percent reside in the Northeast region of the U.S.¹⁰ Over half of the AI/AN population resides in urban areas of the U.S.¹¹

American Indians and Alaska Natives, on average, are younger, have larger households, are poorer, less educated, and are more likely to reside in rural areas than are other population groups.¹² American Indians who reside on reservations have real per capita income and median household income that is only about half the U.S. level. Poverty rates of Indian people are three times the national average and unemployment rates are more than twice the U.S. average.¹³ Similarly, the AI/AN population has long experienced a disproportionately high rate of various health problems.¹⁴

Much of the available data and research on AI/AN health and well-being does not distinguish AI/AN subgroups, although it may sometimes distinguish between those who report AI/AN as their sole race and those who report AI/AN race in combination with other races. The change in racial reporting on the 2000 Census and other federal surveys has led to a dramatic increase in the number of people reporting AI/AN race, due to reporting of multiple races. This may affect how changes in health and well-being are measured for AI/AN between pre-2000 data periods and 2000 and later. Data on AI/AN who are members of federally recognized tribes are, however, available from the Indian Health Service and the Bureau of Indian Affairs, although these are often limited to AI/AN who reside on or near American Indian reservations or Alaska Native villages.

2.1.2 Native Hawaiians and Other Pacific Islanders

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 including the dependent territories of the U.S. (Guam, American Samoa, and the Northern Mariana Islands) and Freely Associated States for which the U.S. provides defense, funding grants, and social services to its citizens (Micronesia, Palau, and the Marshall Islands).

The U.S. Census surveys Hawaii and also identifies those who reside in the 50 states and the District of Columbia who identify race as Native Hawaiian, Guamanian or Chamorro, Samoan, or other Pacific Islander, but does not conduct the Census in the Pacific Insular Areas. Thus, Census data are available only for Native Hawaiians and other Pacific Islanders who reside

¹⁰ U.S. Census Bureau, "The American Indian and Alaska Native Population: 2000." Census 2000 Brief, February 2002.

¹¹ *Ibid.*

¹² *Ibid.*

¹³ Taylor, J.B. and Kali, J.P., "American Indians on Reservations: A Databook of Socioeconomic Change Between the 1990 and 2000 Censuses." The Harvard Project on American Indian Economic Development, Harvard University, Cambridge, MA. January 2005.

¹⁴ Promises to Keep: Public Health Policy for American Indians & Alaska Natives in the 21st Century, M. Dixon and Y. Roubideaux (eds.). The American Public Health Association, 2002; T. Kue Young, The Health of Native Americans: Towards a Biocultural Epidemiology, Oxford University Press, 1994; American Indian Health: Innovations in Health Care, Promotion, and Policy, E.R. Rhoades, M.D. (ed.). The Johns Hopkins University Press, 2000.

in the U.S. Those who identified their race, alone or in combination with another race, as Hawaiian/Pacific Islander on the 2000 Census constituted 0.3 percent of the U.S. population.¹⁵ Only one-third of the 874,414 Native Hawaiians/Pacific Islanders identified in the 2000 Census reported that their race was Native Hawaiian/Pacific Islander alone, while two-thirds reported their race as Hawaiian/Pacific Islander in combination with another race. Nearly three-fourths of Native Hawaiians and other Pacific Islanders resided in the Western U.S. in 2000; over half resided in Hawaii and California. Native Hawaiians and other Pacific Islanders are nearly 25 percent of the population of Hawaii.

Within the Native Hawaiian/Pacific Islander 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/PI. Micronesians, including Guamanians, Marshallese, Palauans, residents of the Northern Mariana Islands and of the Federated States of Micronesia, are 13 percent of all NH/PI. Melanesians, including Fijians, New Caledonians, Solomon Islanders, Vanuatians, and Papua New Guineans, are two percent of this population.¹⁶

There is less information available on health and well-being status of NH/PI than for AI/AN.¹⁷ Until recently, race identification on the U.S. Census and other national surveys combined Asians and Native Hawaiians/Pacific Islanders into one category. There has been increasing evidence, however, that there is great diversity within this combined racial category. As a result, the Office of Management and Budget established a new racial category, Native Hawaiian and Pacific Islander, 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 for this group.¹⁸

The limited information available indicates that the NH/PI population is younger, household size is larger, and household income is lower than the U.S. average. Generally, available data also indicate that the NH/PI population experiences poorer health than the U.S. average. Infant mortality rates for Native Hawaiians in 2002 were 37 percent higher than the rate for all races in the U.S. Furthermore, the tuberculosis rate in the U.S. Pacific Islands was 8.4 times the mainland U.S. average; and Native Hawaiians in Hawaii were more than twice as likely to be diagnosed with diabetes as non-Hispanic white residents of Hawaii.¹⁹

Limited data are available for NH/PI as a combined group, particularly prior to 2000, although the shift to separate identification of NH/PI in federal surveys will lead to increasing availability. However, the availability of information on NH and PI as separate population

¹⁵ U.S. Bureau of the Census, "The Native Hawaiian and Other Pacific Islander Population: 2000." Census 2000 Brief, December 2001.

¹⁶ Panapassa, S.V. "The Health of U.S. Pacific Islander Populations: Emerging Directions." Presentation, May 2005.

¹⁷ National Committee on Vital and Health Statistics, Health Data Needs of the Pacific Insular Areas, Puerto Rico, and the U.S. Virgin Islands, Subcommittee on Populations, December, 1999.

¹⁸ *Ibid.*

¹⁹ U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Office of Minority Health, "Highlights in Minority Health, Asian American and Pacific Islander Heritage Month." May 2005. (accessed 9-16-05 at <http://www.cdc.gov/omh/Highlights/2005/HMary05.htm>).

groups and of subgroups within the PI population is very limited, and this is unlikely to change in the absence of new initiatives. This is particularly the case for Pacific Islanders who reside in Pacific Insular Areas, since even the U.S. Census is not conducted in those areas.

2.2 Findings of Prior Studies and Reports

A number of past studies and reports pertaining to AI/AN/NA data availability, factors that limit data availability, and strategies for improving availability and quality of these data were identified during this project. A brief review of these prior studies and reports is presented in this section. A complete listing of these reports is presented in Appendix C.

2.2.1 Availability and Quality of Data on AI/AN/NA Health and Well-being

In 1999, DHHS/ASPE funded a study of the utility of federal databases for analysis of Hispanic and Asian or Pacific Islander subgroups and Native Americans.²⁰ An inventory was conducted of 14 federal data sets, including assessment of the sample sizes of AI/AN (and Asian/PI as a combined group) and implications of these sample sizes for the ability of the selected data sets to provide data on these subgroups with adequate precision for most practical uses. The study also focused on developing strategies and evaluation of methods that could be used to enhance this ability for surveys with insufficient sample sizes.

Findings of the review of the 14 data sets indicated the following:

- All of the selected data sets identified AI/AN as a combined group and a subset of the 14 data sets identified AI and AN as separate groups;
- Only two of the data sets—Census and the American Community Survey (ACS)—had sample sizes of AI/AN sufficient for reliable estimates. Three data sets had samples large enough to produce limited cross-classifications with acceptable precision levels, and two additional data sets had sample sizes large enough to produce simple distributions;
- Only the Census and ACS had sample sizes for Pacific Islanders that were large enough to produce reliable estimates, with four other data sets having large enough samples to produce simple distribution for this group;
- Only two data sets (Census and ACS) had sufficient samples to produce reliable estimates for Native Hawaiians, with no other data sets having sufficient samples to produce even simple distributions; and
- Seven of the 14 data sets examined did not have large enough samples of AI/AN and NH/PI for even simple distributions to be produced.

²⁰ Waksberg, J., Levine, D., and Marker, D. “Assessment of Major Federal Data Sets for Analyses of Hispanic and Asian or Pacific Islander Subgroups and Native American: Task 3 Report: Extending the Utility of Federal Databases.” Prepared for DHHS/ASPE: Westat, Rockville, MD, May 2000.

Another, more recent, study examined the availability of AI/AN data for measuring health disparities under the methodology developed for the National Healthcare Disparities Report (NHDR) that is submitted annually to Congress.²¹ The methodology for the NHDR relies on 149 indicators of quality (e.g., “rate of cervical cancers diagnosed at late stage,” “current smokers age 18 and over receiving advice to quit smoking”) and 60 indicators of access to care (e.g., “people under age 65 with health insurance,” “people with a dental visit in the past year”) that are compared to assess disparities for racial, ethnic, and socioeconomic groups. Data on quality measures were available for Blacks for all measures, for Hispanics for 95 percent of measures, and for Asians for 63 percent of measures. Data on access measures were available for Blacks and Hispanics for all measures and for Asians on 84 percent of measures. For the AI/AN group, data on health care quality and access measures were available for only 42 percent of these measures, and no data were available for patient safety measures (one of the quality indicators).

A commentary on the Moy *et al.* paper on gaps in the NHDR data measures for the AI/AN population²² suggests that the paper likely underestimates the data problem, due to misclassification of race on some data sets that appeared to be adequate for analysis. The author also conducted a separate review of 190 large or national health data sets to assess availability of data on older AI/AN populations and found that only 13 percent of these data sets contained data on 100 or more AI/AN ages 65 and older.

Another DHHS/ASPE funded study entitled *Rural Research Needs and Data Sources for Selected Human Services Topics* provides additional information on the scarcity of data on well-being measures for populations residing in rural areas—where about half of the AI/AN population resides.²³ Findings of the study identified rural research gaps that could be addressed using state administrative data sources such as: 1.) complete national data on child care subsidies, transportation subsidies, and Workforce Investment Act services are currently unavailable; however, selected states maintain detailed databases on these topics that could be used to examine rural services; and 2.) states provide information on substance abuse treatment and child welfare to the federal government, which removes (or does not collect) detailed geographic identifiers. By working with states to ensure confidentiality and privacy protection, researchers can use detailed geographic identifiers to examine rural facilities and clients. Furthermore, a recent Government Accountability Office study²⁴ of the implementation of the Indian Child Welfare Act (ICWA) similarly reported that it was difficult to draw conclusions about the ICWA as there was no national level database and only five states were found to be able to identify American Indian children who came under the Act’s provisions.

²¹ Moy, E., Smith, C.R., Johansson, P., and Andrews, R., “Gaps in Data for American Indians and Alaska Natives in the National Healthcare Disparities Report.” *American Indian and Alaska Native Mental Health Research: The Journal of the National Center*. American Indian and Alaska Native Programs, University of Colorado at Denver and Health Sciences Center: 13 (1), 52-69. 2006.

²² Rhoades, D. “Commentary – Disparities in Data for American Indians and Alaska Natives,” *American Indian and Alaska Native Mental Health Research: The Journal of the National Center*, American Indian and Alaska Native Programs, University of Colorado at Denver and Health Sciences Center: 13(1) 70-74. 2006.

²³ Strong, D.A., Del Grosso, P., Burwick, A., Jethwani, V., and Ponza, P., “Rural Research Needs and Data Sources for Selected Human Services Topics.” Prepared for U.S. DHHS/ASPE: Mathematica Policy Research, Princeton, NJ, August 2005.

²⁴ U.S. Government Accountability Office, “Report on Implementation of the Indian Child Welfare Act,” Washington, DC: April 8, 2005.

The issue of misclassification of AI/AN race on state death certificates was examined in 1996 by the Indian Health Service.²⁵ The methodology involved matching National Death Index records to records from the patient registration system maintained by IHS, which provides health care to members of federally recognized American Indian tribes and Alaska Native villages. Findings indicated that a race other than AI/AN was reported on the death records for 11 percent of those decedents who were IHS patients. In addition, there was substantial variation among states in the rate of miscoding of AI/AN race, ranging from 28 percent miscoded to 2 percent miscoded.²⁶

2.2.2 Factors Contributing to Gaps in AI/AN/NA Data

A number of the studies cited above and others that were reviewed included discussion of the factors that contribute to gaps in the data on AI/AN/NA health and well-being. Key factors and issues identified in these studies include:

- Small population sizes for AI/AN and NH/PI that result in inadequate sample sizes for analyses in most federal data sets.²⁷
- Concentration of the AI/AN and NH/PI populations in rural areas and native homelands that produce especially small samples.²⁸
- Pacific Islanders residing in Pacific Insular Areas are excluded from many federal data sets.²⁹
- Privacy/confidentiality protection rules that limit use of small samples in geographically identified areas with scarce population.³⁰
- Misclassification of race on vital statistics and other administrative data sets (e.g., disease surveillance databases, disease registries, and the Medicare Enrollment Database).³¹
- Lower response rates to surveys of the AI/AN/NA population, due to language barriers, higher mobility, lack of telephones, and cultural issues.³²

²⁵ Indian Health Service, U.S. Department of Health and Human Services, Adjusting for Miscoding of Indian Race on State Death Certificates. Rockville, MD: 1996

²⁶ The Indian Health Service is currently updating the 1996 study to determine whether there have been changes in the miscoding of AI/AN race on death certificates.

²⁷ Waksberg *et al.*, *op cit.*; Strong *et al.*, *op cit.*; U.S. DHHS Data Council, Meeting Minutes, April-October 2005.

²⁸ Strong *et al.*, *op cit.*; National Committee on Vital and Health Statistics, Subcommittee on Populations. Eliminating Health Disparities: Strengthening Data on Race, Ethnicity, and Primary Language in the United States. Hyattsville, MD: August 2005; National Committee on Vital and Health Statistics, Health Data Needs of the Pacific Insular Areas, Puerto Rico, and the U.S. Virgin Islands, Subcommittee on Populations, December, 1999; National Committee on Vital and Health Statistics, Subcommittee on Populations, Proceedings of Meeting on Asian, Native Hawaiians, and Other Pacific Islanders Data on Race/Ethnicity, May 22, 2003.

²⁹ NCVHS, *Ibid*

³⁰ Strong *et al.*, *op cit.*

³¹ Rhoades, *op cit.*; Moy *et al.*, *op cit.*

³² Finch, B.K., Morton, S., Elliot, M.N., Golinelli, D., Lurie, N., Do, D.P., Rastegar, A., Griffin, A.R., and Valentine, D., Draft Report: Evaluation of Statistical Methods for Data Collection and Analysis on Racial and Ethnic Minorities and Other Hard-to-Reach Populations. Contract No. 282-00-0005. Prepared for U.S. DHHS,

2.2.3 Strategies for Improving AI/AN/NA Data Availability and Quality

Several of the studies and reports that were reviewed included suggestions and strategies that, if implemented, could address some of the factors that contribute to gaps in AI/AN/NA data and improve data availability and quality. The most comprehensive set of strategies presented are those contained in the Joint Report of the DHHS Data Council's Working Group on Racial and Ethnic Data and the Data Work Group for the DHHS Initiative to Eliminate Racial and Ethnic Disparities in Health.³³ These recommendations are listed below by topic area.

Current Surveys and Survey Data. The Joint Report recommended developing a 10-20 year strategic plan for the national surveys that includes a schedule for periodic targeting of racial and ethnic groups for inclusion in the surveys; supporting extramural and intramural analyses of existing data on and related to racial ethnic groups; encouraging minority researchers through training programs and funding; and encouraging formation of partnerships between minority institutions, researchers, and institutions that are expert in these areas. Other recommendations included developing aggressive public use data release programs to promote wider analysis of minority health data and data relating to human services, including financial support of analysis and training; making data available on websites; and supporting periodic data user conferences. In addition, there was support in the report for increasing culturally sensitive analysis and interpretation.

Measurement and Methods. Recommendations related to improving measurement and methods included: ensuring that the standards being developed in response to the Health Insurance Portability and Accountability Act (HIPAA) include racial and ethnic identifiers consistent with the OMB standard classifications, and developing methodological and statistical bridges from old racial/ethnic classifications to the new OMB classification system. In addition, the Joint Report recommended that DHHS assume a leadership role to undertake methodological research about issues associated with measurement of race/ethnicity in surveys, censuses and research and to conduct research to improve reporting of race/ethnicity for administrative and medical records. These recommendations also include conducting research on the effect of survey methods on response rates and new methods for improving response rates and quality of survey data.

New Data Collection. Recommendations of the Joint Report for new data collection include addressing the racial and ethnic data gaps for objectives included in Healthy People 2010 with the initial focus on the six health focus areas pertaining to eliminating disparities; expanding or establishing new registries for certain chronic conditions targeted in the eliminating disparities initiative of Healthy People 2010, including cancer, diabetes, heart disease and stroke; and improving the quality of racial and ethnic data in existing registries. Other recommendations for new data collection include exploring the feasibility of collecting sufficiently large sample sizes

Office of Minority Health, Rockville, MD: Undated.

³³ U.S. DHHS, "Improving the Collection and Use of Racial and Ethnic Data in HHS," Joint Report of the DHHS Data Council's Working Group on Racial and Ethnic Data and the Data Work Group for the DHHS Initiative to Eliminate Racial and Ethnic Disparities in Health (undated).

in each state for each racial and ethnic group that comprises a significant proportion of the population of that state.

Other Recommendations. In addition to these broad recommendations, the Joint Report included a number of more detailed recommendations including: 1) studying the feasibility of collecting geographically identifiable information that would permit geocoding to allow linkages of files in order to fill data gaps; 2) improving and augmenting the data currently available to measure racial discrimination in health care settings; 3) identifying best practices and strategies to address health disparities and racial discrimination; 4) advocating for and requiring inclusion of racial/ethnic data in administrative records for all health and non-health DHHS programs; 5) developing a strategy for standards for electronic patient records; 6) expanding health databases for Puerto Rico, U.S. Virgin Islands, and Pacific Insular Areas; 7) publishing periodic national reports on the health and services received by members of racial/ethnic groups, compared to all races and the White population; 8) developing an initiative to train health personnel in completing records, statistics, survey research, and epidemiology in order to improve racial/ethnic data reporting; 9) strengthening and expanding cooperative efforts with other personnel (e.g., registrars, funeral directors) to improve racial/ethnic identification on vital statistics records and registries; and 10) disseminating relevant racial/ethnic data and findings back to the communities where the data were collected.

These recommendations for improvement of racial/ethnic data are general to all racial/ethnic subgroups of the population and would lead to improvements in the availability and quality of AI/AN/NA data on health and well-being as well as for other groups.

Several studies and reports reviewed focused specifically on strategies for improving AI/AN and NH/PI data. Many of these suggestions for improving data for these groups were mentioned in more than one study or report and are presented below with footnotes indicating the specific sources of the suggestions. These included:

- Aggregate multiple years of data for data sets with insufficient annual samples of AI/AN and NH/PI.³⁴
- Oversampling AI/AN and NH/PI in federal surveys, on a rotating basis, permanently, or through special data collection initiatives, and develop sampling strategies that would be more cost-effective to reach small populations (e.g., cluster sampling).³⁵
- Conduct more methodological research to develop statistical strategies for making maximum use of small population samples (e.g., using small area estimation techniques), while protecting privacy and confidentiality; improve response rates by addressing language, high mobility, and cultural issues; and improve quality of data through studies of cultural differences in interpretation of questions.³⁶

³⁴ Waksberg, *et al.*, *op cit.*; Finch, *et al.*, *op cit.*; U. S. DHHS Data Council, Meeting Minutes, April-October 2005. See, for example, P.M. Barnes, P.F. Adams, and E Powell-Griner, "Health Characteristics of the American Indian and Alaska Native Adult Population: United States, 1999-2003," *Advance Data from Vital and Health Statistics*, Number 356, April 27, 2005, for an example of this approach.

³⁵ Waksberg, *et al.*, *op cit.*; Finch, *et al.*, *op cit.*; Strong *et al.*, *op cit.*; U. S. DHHS Data Council, Meeting Minutes, April-October 2005.

³⁶ Finch, *et al.*, *op cit.*; U. S. DHHS Data Council, Meeting Minutes, April-October 2005.

- Improve accuracy of racial identification on vital statistics and administrative databases through establishing standardized protocols and providing training for those who report vital statistics within states and local areas.³⁷
- Develop protocols for drawing samples of self-reported AI/AN/NA from the American Community Survey conducted by the U.S. Census for use in cost-effectively identifying AI/AN/NA for inclusion in DHHS surveys and to permit linkage of data from multiple surveys with ACS socio-economic and other data.³⁸
- Develop standardized protocols for surveys and provide training to AI/AN/NA communities, states, and research organizations to: 1) assist them to collect community-level data comparable to those collected through federal surveys with insufficient sample sizes to permit analysis of AI/AN and NH/PI health and well-being, 2) augment federal data collection and 3) build data sets for subgroups within the AI/AN/NA population.³⁹

Several of the studies and reports reviewed also identified a need for data on: specific subgroups of the AI/AN and NH/PI populations, including data on federally recognized AI/AN tribes, due to differences among tribes and geographic regions in health and well-being; AI/AN residing in urban areas; Native Hawaiians and Pacific Islanders separately; and subgroups of the Pacific Islander population.

3 METHODS

The objectives of this report are to: 1) identify gaps in the availability and quality of AI/AN/NA data on health and well-being; 2) identify strategies for data improvement that may be considered by DHHS and others; and 3) identify current initiatives within DHHS and other federal agencies that are intended to improve the availability and quality of AI/AN/NA data. The methodology used to address these objectives included:

1. Review of selected literature and published and unpublished reports.
2. Analysis of the *Data Catalog on AI/AN/NA Health and Well-being*.
3. Key informant interviews with a number of DHHS staff and other knowledgeable individuals.

Each of these analysis and data collection methods is described in this section.

³⁷ Rhoades, *op cit.*; National Committee on Vital and Health Statistics, Subcommittee on Populations. Eliminating Health Disparities: Strengthening Data on Race, Ethnicity, and Primary Language in the United States. Hyattsville, MD: August 2005; National Committee on Vital and Health Statistics, Subcommittee on Populations, Minutes of Meeting on AI/AN Issues. Denver, CO, September 27, 2002.

³⁸ Waksberg *et al.*, *op cit.*

³⁹ National Committee on Vital and Health Statistics, Subcommittee on Populations. Eliminating Health Disparities: Strengthening Data on Race, Ethnicity, and Primary Language in the United States. Hyattsville, MD: August 2005; U.S. DHHS, NCVHS, Subcommittee on Population – AI/AN Issues. Denver, CO, September 27, 2002.

3.1 Review of Selected Literature and Reports

The literature and published and unpublished reports reviewed for this report were obtained from several sources. Representative from the DHHS Data Council provided a substantial quantity of published and unpublished reports on racial/ethnic data availability and quality and strategies for improving data at the initial project meeting. Key informant interviews provided information on additional studies that were then obtained, and some members of the DHHS Data Council forwarded additional references to us throughout the project. We also conducted a search of the literature. Each paper and report was reviewed for information on data availability and quality and gaps, information on factors contributing to gaps, suggestions/strategies for improving data availability and quality, and examples of current initiatives and/or application of specific strategies.

3.2 Analysis of Data Catalog Information

The *Data Catalog on AI/AN/NA Health and Well-being*, an earlier product of this study, provided the foundation for the assessment of the availability and quality of data on AI/AN/NA health and well-being and the identification of gaps in these data. As a first step in the development of the data catalog, we identified a set of policy areas relevant to AI/AN/NA health and well-being, with input and review by DHHS. These key policy areas include health policy⁴⁰ and well-being indicators including: child well-being, economic well-being, education levels and opportunities, elder well-being, family well-being, housing indicators, justice system indicators, military/veterans' issues, and transportation quality and availability.

A detailed list of the indicators under each of these policy areas is provided in Appendix A to this report. Each data source was reviewed and classified as providing or not providing data appropriate for examining each of these policy areas/issues.

The approach used to compile the data catalog included identification of an initial list of potential data sources through a comprehensive review of federal agency sources, review of data repositories, consultation with members of the DHHS Data Council's Racial and Ethnic Data Working Group, and consultation with project consultants and a small AI/AN/NA workgroup composed of stakeholders and technical experts from AI/AN and Native Hawaiian communities. A total of 153 possible data sources were identified. Because project resources were limited, these 153 data sources were assigned priorities for review by the project staff and the ASPE Task Order Officers, based on the policy areas covered and the specific population groups included.

⁴⁰ Health policy topics considered in the catalog include such areas as health status (e.g., self-reported health status, disability rates, mortality/morbidity rates, trends over time); disease-specific prevalence and incidence (e.g., prevalence of diabetes, tuberculosis, sexually transmitted diseases); 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, utilization rates, health insurance coverage, health care financing, socioeconomic factors, preventative measures); evidence-based practices and programs that address causes of health disparities, result in positive health outcomes, and are generalizable/replicable; and the role of traditional medicine in AI/AN/NA communities.

Of the initial 153, 111 were identified as highest priority for preliminary screening. Of these 111 that were subjected to preliminary screening, a total of 67 were identified as having data that met specific technical parameters, identified AI/AN and/or NH/PI race and were likely to have sufficient sample size to produce data of utility to researchers and policymakers. These 67 data sources were reviewed in depth, using a detailed protocol, and profiles of each data source were prepared describing up to 25 relevant elements of the data set.

The detailed profiles of these 67⁴¹ data sets were used for the analysis in this report to identify gaps in availability of data. The data catalog analysis focused on: 1) identification of primary policy areas for which there are no available data for the AI/AN, AI only, AN only, tribal affiliation, NH/PI combined, NH only, and PI only populations; 2) identification of geographic levels of analysis for each group that were possible using these data sets; 3) identification of the number of data sets available on AI/AN, AI only, AN only, NH/PI, NH only, and PI only populations, for each policy area; and 4) identification of the number that have sufficient sample size (200 or more) to permit analysis to be conducted without necessarily requiring aggregation of multiple years of data.

Although the criterion for sample size was set at a minimum of 100 in the data catalog, this sample size is sufficient only for very limited descriptive analysis of the health and well-being characteristics of the entire population. We chose a minimum sample size of 200 for the gaps analysis because a sample of this size or greater would likely produce reasonably precise estimates for analyses of key issues in some subgroups of the AI/AN and NH/PI population (e.g., males compared to females, comparison of broad age groups).⁴²

3.3 Key Informant Interviews

Informal telephone discussions were held with 13 knowledgeable individuals to obtain additional information for this report. Several of the members of the DHHS Data Council's Work Group on Racial and Ethnic Data suggested names of people that they thought were particularly knowledgeable about AI/AN/NA data sets or current initiatives. These individuals were contacted, interviewed, and also asked for recommendations of other people who might be useful sources of additional information. Ten of the contacts were within DHHS, one was from the Census Bureau, and two were not employed by the federal government but had been involved through prior work pertaining to AI/AN/NA research or data. Contacts were asked about their perceptions of data gaps, changes being made in data collections, and suggestions for improvements, etc. This material informs the remainder of this report.

⁴¹ The data catalog contains 68 different profiles. However, two of those profiles (Census 2000 and the Census 2000 – American Indian and Alaska Native Summary File) pertain to a single data set. For this paper, only Census 2000 was included in the analysis.

⁴² Waksberg et al. determined the effective sample sizes required to produce estimates of specified precision (i.e., margin of error) when the variable of interest has one characteristic with a prevalence of 15 to 20 percent. They indicate that sample sizes ranging from 100 to 400 are required to achieve estimates with precision levels of .30 to .10 in these types of variables.

3.4 Limitations

There are several limitations of this study that should be considered by users of the information presented in the findings section below. First, users should be aware that the data catalog that was the primary source of information on availability of AI/AN/NA health and well-being data does not include the universe of all public and private data sets that include identifiers for AI/AN/NA race. However, our process for identifying publicly available data sets, including discussion and input from the DHHS Data Council and other DHHS staff, our working group members, and review of published and unpublished reports, produced 153 databases that were initially reviewed, 111 data sources that were screened, and 67 that received complete detailed review. As a result of this process and multiple reviewers of data catalog, we believe it is unlikely that there are many other federal or publicly available databases that have sufficient sample sizes of AI/AN/NA, include measures of health and well-being, and meet the technical criteria established for the data catalog. (There are of course still yet fewer data sets containing information on individual tribes or other subgroups.)

Another limitation may exist due to the fact that the search for data sources was based upon a set of AI/AN/NA major health and well-being policy areas that was developed to guide the search (Appendix A). Our search, therefore, produced data sets that included information on these specific policy areas. To the extent that there are other policy areas that are relevant to AI/AN/NA health and well-being that were not included in the search, this report does not provide information on the availability of data to examine those issues. In addition, the review of data sources included in the data catalog did not include an in-depth examination of the quality of the AI/AN/NA data in each database. While the sample size of AI/AN/NA respondents was identified for each database, it was not possible within the existing resources to examine the extent to which there may be missing data for specific policy-related questions or other issues that might affect the usability or reliability of the data. Some data sets that appear to provide data on specific policy areas for AI/AN/NA population groups may not be adequate for all research and policy uses. As a result, it is possible that there are more gaps in AI/AN/NA data availability than are reflected in this report.

Finally, it is possible that this report understates the number of current and planned initiatives underway within DHHS and other federal agencies that may result in improved AI/AN/NA health and well-being data. Key informant interviews were conducted with knowledgeable individuals on this issue, but only 13 interviews were conducted and the majority of these were with DHHS staff. There may well be current or planned initiatives within DHHS and other federal agencies of which these interviewees were unaware.

4 SUMMARY OF DATA AVAILABILITY AND GAPS IN SELECTED DATA SETS, BY POLICY AREA AND POPULATION GROUP

The detailed profiles on 67 data sets in the data catalog were analyzed to assess policy areas for which data were available, for which there were limited sources of data, and for which

there were gaps in the data. The analysis was conducted for each of the population groups that are the focus of this study, American Indians/Alaska Natives, Native Hawaiians/Pacific Islanders combined, as well as American Indians, Alaska Natives, Native Hawaiians and Pacific Islanders as separate groups. In this section, we review data availability by policy area and population group, discuss the availability of geographic data for these policy areas and population groups, and identify policy areas for which there are gaps in data available or limited data available for each population group. It is important for the reader to keep in mind that, when data are identified as “available” and when we report “no gaps” in the data, this does not necessarily mean that there are sufficient and available data for all measures of health and well-being of the AI/AN/NA population overall or for specific subgroups. However, these findings do indicate that some data exist that will permit measurement of health or well-being for these population groups.

4.1 Data Availability, by Policy Area and Population Group

Data on a specific policy issue, for each of the population groups, was deemed to be available if there was at least one data set that focused primarily on the policy issue and that included sample members that were identifiable as a member of the population group. It is important to restate here that the search for data sets was guided by the set of policy issues that was developed in the initial stages of this study; consequently, the information on which this analysis is based consists of a purposively chosen subset of all possible data sets. The initial information on availability of data, by policy area and for the combined population groups, is presented in Table 1 below. Although there are 67 data sets in the catalog, a single data set may cover multiple policy areas. Data sets are considered to “cover” a policy area if there are a substantial number of meaningful survey items concerning that policy area. Data sets that contain only one or two questions in an area are not included.

Table 1. Number of Data Sets that Identify Specific Population Groups, by Policy Area

Policy Issue	Number of Data Sets	
	Covering Area	
	AI/AN	NH/PI
Child Well-being	16	9
Economic Well-being	15	10
Education	11	8
Elder Well-being	8	4
Family Well-being	14	9
Health Policy Issues	37	20
Housing	8	5
Justice System Issues	9	5
Military Service/Veterans Issues	2	1
Transportation	3	2

Note: Sixty-seven total data sets were reviewed. Many included data on measures that could be used to examine more than one health and well-being policy area and, therefore, were included more than once in the table.

Data availability findings, by policy area and for each population group, include:

- Data are available for the AI/AN population for all of the policy areas. However, the number of available data sets varies by policy areas. The policy area with the least number of available data sets for the AI/AN population is military/veterans issues, with only two data sets identified.
- Data are available for examination of all of the 10 policy areas for the combined NH/PI group. However, for this group, there are fewer data sets available that provide information on most policy areas than for the AI/AN population. Additionally, there are two policy areas for which there are only one or two data sets available (i.e., military/veterans issues and transportation). This represents a potential gap in the available data, since it is possible that these one or two data sets provide the potential for limited examination of the policy issue.

A related issue is the availability of data for subgroups of the populations that are the focus of this study. For American Indians and Alaska Natives, for example, it would be useful for policy reasons to have identifiable data on American Indians and Alaska Natives separately and on members of federally recognized tribes, both those who live on/near reservations and those that live off-reservation. Details on the distribution of all subgroup identifiers by policy area are depicted in Table 2 below.

Table 2. Number of Data Sets that Identify Specific Subpopulation Groups

Policy Issue	AI/AN	AI	AN	NH/PI	NH	PI
Child Well-being	16	0	0	9	3	2
Economic Well-being	14	3	3	10	3	3
Education	11	2	2	8	3	3
Elder Well-being	8	1	1	4	2	2
Family Well-being	14	2	2	9	3	3
Health Policy Issues	36	4	4	20	9	5
Housing	8	2	2	5	2	2
Justice System Issues	6	0	0	5	0	0
Military Service/Veterans Issues	2	0	0	1	1	1
Transportation	3	1	1	2	1	1

Note: Sixty-seven total data sets were reviewed. Many included data on measures that could be used to examine more than one health and well-being policy area and, therefore, were included more than once in the table.

Seven of the 10 policy areas have data sets that identify American Indians and Alaska Natives as separate groups. The three policy areas that do not include data sets with these identifiers are child well-being, justice systems issues, and military/veterans issues. There are more data sets that identify Native Hawaiians and Pacific Islanders as separate groups than the AI and AN groups. The only policy area that does not identify these groups separately is the justice system issues area.

The potential for examining health and well-being of AI/AN and NH/PI population subgroups is less than for the combined AI/AN and combined NH/PI population groups. The

majority of available data sets report race only for the combined groups. Only 5 of the 67 data sets profiled permit separate analysis of American Indians and Alaska Natives, while 12 permit separate analysis of Native Hawaiians and 8 permit analysis of Pacific Islanders separately from Native Hawaiians. To the extent that there are differences among subgroups of the AI/AN and NH/PI populations, it may be important for measurement and monitoring of disparities and for research to be able to distinguish among these subgroups. However, this is not possible for some of the policy issues.

4.2 Sample Sizes 200 or Greater, by Policy Area and Population Group

A second level of analysis of availability was then conducted to determine the number of data sets, for specific policy issues and each population, that had a sample size for the relevant population of at least 200. As noted earlier, the sample size criterion was set at 100 for a data set to be included in the data catalog. Analyses that are possible with a sample size of 100, however, are limited to description of characteristics of the population. For this report on availability and gaps in data, we chose to highlight data sets of sample size 200 or more. A sample size of at least 200 offers the possibility of producing reliable and valid information on specific issues, although larger sample sizes would be necessary to permit analysis of the specific issue for more than one or two characteristics of sample members. Data sets with fewer than 200 sample size may require aggregation of multiple years of data and/or other statistical strategies for useful analysis. Of the 67 data sets reviewed, 50 had 200 or more persons in the AI/AN combined category and 20 had 200 or more persons in the NH/PI combined category. Only 3 data sources had 200 or more persons in the separate categories for AI and AN, only 4 data sources had 200 or more in the NH category and only 3 data sources had 200 or more in the PI category. Table 3 displays the number of data sets that have a sample size of 200 or more for each of the specific population groups.

Table 3. Number of Data Sets with Sample Size 200 or More by Population Group

Policy Issue	AI/AN	AI	AN	NH/PI	NH	PI
Child Well-being	14	0	0	6	0	0
Economic Well-being	11	1	1	5	2	2
Education	9	1	1	2	1	1
Elder Well-being	6	0	0	2	0	0
Family Well-being	13	1	1	4	2	2
Health Policy Issues	29	3	3	12	3	2
Housing	6	1	1	3	2	2
Justice System Issues	4	0	0	2	0	0
Military Service/Veterans Issues	2	0	0	0	0	0
Transportation	3	1	1	2	1	1

Note: Sixty-seven total data sets were reviewed. Many included data on measures that could be used to examine more than one health and well-being policy area and, therefore, were included more than once in the table.

When the available data sets are examined for sample sizes of at least 200, the number of data sets available to examine specific policy issues is reduced. For most policy issues and

population groups, however, there are data sets that can be used to produce information without requiring aggregation of data sets or special statistical techniques. Findings include:

- For the combined AI/AN population, there are at least two data sets with a sample size of 200 or greater for each policy area.⁴³
- For the combined NH/PI population, nine of the 10 policy areas contain data sets with a sample size of 200 or greater. Only one policy area, military service, does not.
- For the AN, AI, NH and PI populations separately, there are fewer policy areas that have sample sizes of 200 or greater. Policy areas that meet this criterion include economic well-being, education, family well-being, health policy issues, housing and transportation.

Detailed information on sample sizes of AI/AN, NH/PI, AI, AN, NH and PI separately for each of the 67 data sets, organized by policy area, are provided in Appendix B to this report.

It is not surprising that there are few data sets in most policy areas that have sample sizes of 200 or more for these populations. The AI/AN population is only 1.5 percent and the NH/PI population is only 0.3 percent of the U.S. population and, as a result, is unlikely to be well-represented in most national surveys. For some ongoing surveys, it may be possible to aggregate multiple years of survey data to increase sample sizes. For other surveys, however, increasing the availability of data that permit detailed analysis of health and well-being of these groups would require strategies that focus on over-sampling or conducting surveys focused specifically on these populations.

4.3 Geographic Information Available, by Primary Policy Area and Population Group

The AI/AN population is geographically concentrated in the western U.S. and three-fourths of the NH/PI population reside in the western region of the U.S. Nearly half of AI/AN reside on reservations that were established under treaties between the federal government and tribes or in Alaska Native villages; these reservations and villages are typically located in rural areas. However, over half of AI/AN reside in urban areas and may experience different access to and needs for services to support their health and well-being than the rural AI/AN population. About 15 percent of Native Hawaiians live on Hawaiian homelands held in trust by the State of Hawaii for the benefit of Native Hawaiians. Because of this concentration of the AI/AN and NH populations on reservations and homelands, there is potential policy interest in having data available that permits examination of health and well-being for those residing on reservations or homelands relative to the health and well-being of AI/AN and NH who reside in other areas of the U.S. In addition, substantial differences in health and well-being have been identified within the AI/AN population, by geographic region.⁴⁴ It is possible that similar variations exist for the

⁴³ One elder well-being data set and two education data sets report on 200 or more, but provide data at the program level only. Analysis at the individual level using these data sets is not possible.

⁴⁴ See, for example, U.S. DHHS, Indian Health Service, Regional Differences in Indian Health, Rockville, MD:

NH/PI population, by geographic location. To permit measurement and monitoring of within-group difference, it would be very useful to identify data sets that include geographic identifiers that would enable comparisons of health and well-being of AI/AN and NH/PI populations by location (e.g., U.S. Census Region, IHS Areas, urban and rural).

We examined the available data sets to determine the geographic indicators available on each that might permit analysis of the AI/AN and NH/PI population, by geographic sub-areas. Of the 67 data sets included in the study, 62 are derived from national-level surveys, 3 from data collected in a single state, and 2 data sets were collected in AI/AN communities only. Findings for the AI/AN population indicate that four of these data sets do not include any geographic indicators and can only be used to produce national-level estimates. For eight other data sets, geographic indicators are available for the four Census Regions (i.e., North, South, Central, and West). Seventeen data sets include state indicators and the remaining 36 data sets include indicators for urban and rural areas, counties, and reservations/homelands. The remaining two data sets were collected in AI/AN communities only, as stated earlier.

Although the majority of the data sets include geographic indicators that would permit analysis below the national level, the sample sizes of AI/AN and NH/PI in most of these data sets are small and likely would not be sufficient to permit disaggregation of the AI/AN and NH/PI data to produce reliable estimates at sub-national levels. The exceptions are those data sets that are censuses of the population (e.g., Census 2000, vital statistics records, IHS Resource and Patient Management System data). We did not have sufficient resources to investigate in detail each data set to determine whether specific sub-national analyses could be supported by the sample size available for the AI/AN and NH/PI population and subgroups.

4.4 Data Gaps by Policy Area and Population Group

Our review identifies a number of gaps in data available for examination of AI/AN/NA health and well-being issues. Primary policy areas for which no or limited data are available for specific population groups are shown in Table 4. Policy areas for which no data sets were identified for a population group are indicated as a “Gap.” Policy areas for which only one or two data sets were identified for a population group are indicated as “Limited” availability.

Table 4. Data Gaps, by Policy Area and Population Group–All Data Sets

Policy Area	AI/AN	AI	AN	NH/PI	NH	PI
Child Well-being	>2	Gap	Gap	>2	>2	Limited
Economic Well-being	>2	>2	>2	>2	>2	>2
Education	>2	Limited	Limited	>2	>2	>2
Elder Well-being	>2	Limited	Limited	>2	Limited	Limited
Family Well-being	>2	Limited	Limited	>2	>2	>2
Health	>2	>2	>2	>2	>2	>2
Housing	>2	Limited	Limited	>2	Limited	Limited
Justice System	>2	Gap	Gap	>2	Gap	Gap
Military/Veterans	Limited	Gap	Gap	Limited	Limited	Limited
Transportation	>2	Limited	Limited	Limited	Limited	Limited

Note: The notation “>2” indicates that there are more than two data sources that permit analysis of the policy area for the specific population group. “Gap” indicates that no data set was identified for that policy area for the specific population group. “Limited” indicates that only one or two data sets were identified for the policy area for the specific population group.

There are no policy areas for which there are no data currently available for the AI/AN population and the combined NH/PI population. However, limited AI/AN data are available for military/veterans issues, while limited NH/PI data are available for this issue and for transportation. For AI and AN, as separate population groups, only economic well-being and health policy issues are fully covered by the available data sets; all other policy areas exhibit gaps or limited data for analysis. The separate NH and PI subpopulations have gaps in justice system issues and limited data on child well-being (PI), elder well-being, housing, military/veterans issues, and transportation issues.

Examination of the available data sets with at least 200 sample members for each population group (Table 5) shows that, for the AI/AN combined population there are no gaps, but data are limited for military/veterans issues. Fewer data sources are available for the NH/PI combined group than for the AI/AN combined group, with gaps in available data for military/veterans issues and limited data for four additional policy areas. For the AI, AN, and NH subgroups, only the health policy area has more than two available data sets. The remaining policy areas either have gaps or limited data. When we examine the PI subgroup separately, there are gaps in data availability for four of the policy areas and limited data available for the remaining policy areas.

**Table 5. Data Gaps, by Policy Area and Population Group–
Data Sets With 200 or More Sample Size for Each Population Group**

Policy Area	AI/AN	AI	AN	NH/PI	NH	PI
Child Well-being	>2	Gap	Gap	>2	Gap	Gap
Economic Well-being	>2	Limited	Limited	>2	Limited	Limited
Education	>2	Limited	Limited	Limited	Limited	Limited
Elder Well-being	>2	Gap	Gap	Limited	Gap	Gap
Family Well-being	>2	Limited	Limited	>2	Limited	Limited
Health	>2	>2	>2	>2	>2	Limited
Housing	>2	Limited	Limited	>2	Limited	Limited
Justice System	>2	Gap	Gap	Limited	Limited	Gap
Military/Veterans	Limited	Gap	Gap	Gap	Gap	Gap
Transportation	>2	Limited	Limited	Limited	Limited	Limited

Note: The notation “>2” indicates that there were more than two data sources that permit analysis of the policy area for the specific population group. “Gap” indicates that no data set was identified for that policy area for the specific population group. “Limited” indicates that only one or two data sets were identified for the policy area for the specific population group.

4.5 Discussion

As noted earlier, these analyses suggest that there are no complete data gaps for the combined AI/AN and NH/PI populations. When all data sets are examined for the NH and PI subgroups, there is one gap for each of these subgroups. There are more substantial data gaps for the AI and AN subgroups, with gaps or limited data for all policy areas for these subgroups, with the exception of health and economic well-being.

When the analysis of data availability is restricted to data sources with at least 200 sample members, there are no major gaps in the combined AI/AN group. However, a gap in NH/PI data for military/veterans issues is identified and limited data are available for education, elder well-being, justice system, and transportation. Furthermore, this analysis identifies gaps or limited data for all policy areas for the PI subgroup, while the AI, AN, and NH subgroups have only one policy area (health) that has more than two available data sets.

It is important, also, to remember that each of the examined data sets varies in the specific variables available to examine specific issues within a policy area. For some policy areas and issues, there may be inadequate data for analysis of disparities in health and well-being for these population groups, by demographic, geographic, or economic characteristics, even though some data on the issue/policy area may be collected and usable for aggregate analysis. For example, Census 2000 includes questions on demographic and economic characteristics of respondents, but collects very limited data on health, which precludes using the data set for most health policy issues. The Consumer Assessment of Healthcare Providers and Systems (CAHPS) data sets collect self-reported data on health status and utilization, but do not include questions on income levels which would permit analysis of differences in these measures by economic status.

Furthermore, some of the available data sets are national in scope and the sample size may be adequate to examine policy areas/issues at the national level, but not sufficiently large to permit analysis at the state, region, rural/urban, or reservation/homelands level. Others may be regional and permit examination of the policy/issues only for a geographic subset of the population. For data sets with small sample sizes, aggregation of multiple years may be possible –if the data are collected at frequent intervals and if there are not frequent major changes in the questionnaire or survey methods. These and other methodological issues may have a significant impact on the availability and quality of these data sets for measurement, monitoring, and research on AI/AN/NA health and well-being and, consequently, there may be more (or fewer) data gaps and limitations than are evident in this review.

5 STRATEGIES FOR IMPROVING DATA ON AI/AN HEALTH AND WELL-BEING

This section reviews the issues that affect data availability and quality for AI/AN/NA health and well-being and presents a set of strategies that have been identified that would address a number of the issues that contribute to improving gaps in the data. Information on issues and strategies has been drawn from the literature reviewed, from the discussions held with DHHS and other knowledgeable individuals, and from our own assessment of the information obtained and reviewed for this study. The section concludes with a review of current and planned initiatives within DHHS and other federal agencies that may, over the next few years, result in improved data and greater data availability to address a range of health and well-being issues for the AI/AN/NA population groups.

5.1 Issues that Contribute to Gaps in AI/AN/NA Data Availability and Quality

Several major issues have been identified in the literature and were raised by most interviewees that affect the availability and quality of AI/AN/NA data. These include:

1. Small Population Size. Individuals who self-identify race as only AI/AN are 0.9 percent of the population and those who self-identify as AI/AN in combination with one or more other races are 0.6 percent of the U.S. population. Native Hawaiians and other Pacific Islanders alone or in combination with other races constituted about 0.3 percent of the U.S. population in the 2000 Census. Few national surveys have sufficiently large sample sizes to permit the sample sizes for either of these population groups to be large enough for analysis that would produce reliable and valid estimates of the measures of interest.
2. Geographic Dispersion and Rural Concentration. Much of the AI/AN population is concentrated in rural areas of the U.S. mainland, and there are concentrations of NH/PI in their homelands. Some national surveys under-represent rural areas in their sample designs and this additionally affects the availability of data. Many of the

AI/AN and NH/PI who do not live on reservations or NH homelands reside in urban areas. Unlike some other racial/ethnic groups, the AI/AN and NH/PI populations tend to be dispersed rather than clustered within urban areas.⁴⁵ As a result, it is difficult and costly to develop sample frames that would be appropriate to represent these urban-located populations in national surveys.

3. Misclassification of Race. Vital statistics data collection, particularly vital statistics data on mortality, may misclassify race. Funeral directors and others who complete death certificates may “guess” at the race of the deceased rather than ask a family member. As a result, there may be under-reporting of mortality rates for AI/AN and NH/PI and this under-reporting may vary geographically. Similarly, in some types of administrative databases, intake workers and other staff may not have accurate information on the race of clients and may record their perceptions of race of the client.
4. Lack of/Inconsistent Collection of Race Identifiers in Some Data Sources. There are a number of administrative databases that do not routinely collect data on race/ethnicity or that do not collect these data in a manner that is comparable to the rules set forth in OMB Directive 15. An example is the Social Security Administration (SSA) data, which has not over the years since its inception collected data on race or ethnicity of those who apply for and receive Social Security numbers. Since the Medicare Enrollment Database is created based on SSA eligibility records, the Medicare program does not have accurate and complete race/ethnicity data; however, this program has initiated a number of efforts to collect these data from beneficiaries once they become eligible for Medicare. Federal-state jointly financed programs (e.g., Medicaid) rely on state reporting of race/ethnicity, but states may be inconsistent in their methods of collecting and reporting these data and in the race/ethnicity categories reported.
5. Inadequate Racial Representation, Limited Response Rates, and Question Interpretation. The AI/AN and NH/PI populations are less likely to have telephone service and to have higher mobility than the majority population. As a consequence, they may be excluded from sample frames that rely on telephone numbers or street addresses for sample selection. Similarly, the lack of telephone or mail address may lead to lower response rates due to inability of the survey staff to reach them, when they are included in the survey sample. Other issues that contribute to lower response rates are language and literacy barriers, inadequate efforts to communicate to community leaders and members the value of the survey data to the community, and historical distrust of the federal government. There is also some evidence that differences in culture may be associated with different interpretations of survey

⁴⁵ Finch, B.K., Morton, S., Elliot, M.N., Golinelli, D., Lurie, N., Do, D.P., Rastegar, A., Griffin, A.R., and Valentine, D., Draft Report: Evaluation of Statistical Methods for Data Collection and Analysis on Racial and Ethnic Minorities and Other Hard-to-Reach Populations. Contract No. 282-00-0005. Prepared for U.S. DHHS, Office of Minority Health, Rockville, MD: Undated; Waksberg, J., Levine, D., and Marker, D. “Assessment of Major Federal Data Sets for Analyses of Hispanic and Asian or Pacific Islander Subgroups and Native American: Task 3 Report: Extending the Utility of Federal Databases.” Prepared for DHHS/ASPE: Westat, Rockville, MD, May 2000.

questions, resulting in differences in responses between those from racial/ethnic minority groups and those from racial/ethnic majority groups.

6. Exclusion of Pacific Insular Areas from Some Data Collection. Due to costs and geographic remoteness, some federal and other surveys are not conducted in those Pacific Islands that are U.S. territories or Freely Associated States.
7. Inadequate Collection of Data on AI/AN/NA Subgroups. Several interviewees and authors of research studies and reports indicated that improving AI/AN and NH/PI data was important, but that an issue of equal or greater importance is obtaining data on specific subgroups of these populations. For the AI/AN population, it was suggested that data on members of federally recognized tribes residing on reservations and in urban areas was particularly important, due to the treaty obligations of the federal government to these tribes. For the NH/PI population, it was suggested that it would be very useful to obtain data on NH separately and on specific subgroups of the Pacific Islander population including PI residing in Pacific Insular Areas, including the Freely Associated States.

5.2 Strategies for Improving Data on AI/AN/NA Health and Well-being

The review of selected literature and reports and interviews with knowledgeable individuals identified a number of strategies for improving data availability and quality for AI/AN/NA health and well-being. These strategies are organized and discussed below by the specific type of issue that they address.

Issue 1: Small Population Size

Strategy 1.a: Increase sample sizes of federal surveys to ensure that sufficient samples are obtained to permit analysis of AI/AN and NH/PI health and well-being issues.

The strength of this strategy is that sufficient data would be available from federal surveys on an ongoing basis for measurement, monitoring, and research on AI/AN and NH/PI health and well-being, relative to other racial/ethnic population groups. The National Health Interview Survey (NHIS), for example, currently does not have a sufficiently large sample to permit estimates to be made for the AI/AN or NH/PI population, except through aggregation of multiple years of data. Increasing the overall sample size for the NHIS to ensure that the sample sizes for these populations would be great enough to permit analyses, however, would be very costly to implement.

Strategy 1.b: Over-sample the AI/AN and NH/PI population in federal surveys in order to obtain sufficient sample sizes for these groups.

A subset of this strategy is to design a system that employs rotating over-sampling for specific racial/ethnic groups over a 5-year or 10-year timeframe. This approach would involve

over-sampling of a specific racial/ethnic group once every 5 to 10 years (e.g., Blacks in Year 1, Hispanics in Year 2, Asians in Year 3, AI/AN in Year 4, NH/PI in Year 5, Blacks in Year 6, Hispanics in Year 7, etc., on a planned continuous basis). This latter strategy would ensure that data for these population groups are collected at least once every 5 to 10 years to permit measurement, monitoring, and research on health and well-being disparities. The limitations of this approach primarily involve costs. Over-sampling of these populations for federal surveys, on a routine basis, would increase costs substantially. A rotating over-sampling strategy would be less costly (and, in fact, is being used in some surveys to obtain supplemental data on Blacks, Hispanics, and Asians), but would limit the usefulness of the data for those health and well-being issues that may change cyclically or rapidly (e.g., unemployment, poverty, homelessness, crime patterns). As an example, if there was interest in employment/unemployment rates of AI/AN and data were collected only once every five years, it is possible that the specific years of data collection would coincide with economic cycles of high unemployment. While the estimates could be adjusted statistically, it would be difficult to be certain that these statistical adjustments accurately reflect the actual employment experience of the AI/AN population in the years for which data are not collected.

Strategy 1.c: Aggregation of multiple years of data from surveys that are conducted annually to obtain sufficient AI/AN and NH/PI sample sizes.

Aggregation of multiple years of data is feasible for some, but not all, surveys and this strategy is being implemented within some federal agencies. An example of a federal data set that could be used to analyze AI/AN and NH/PI health and well-being through aggregation of multiple years is the Department of Labor/Bureau of Labor Statistics Consumer Expenditure Surveys (CES), which are available for 1984-2004. In 2004, the CES reported 118 NH/PI in the sample; aggregation of multiple years of CES data, however, would be possible to increase the sample size and permit analysis of a wider range of issues at a higher level of precision. The advantage of this strategy is that it provides a cost-effective way to increase data availability for these population groups, requiring only the time of researchers to combine several years of data and identify and resolve any weighting or other statistical issues that must be addressed. One limitation of this approach is that it may not be appropriate to aggregate the data when the issues to be addressed by analysis are ones that may change rapidly or cyclically. An additional limitation of aggregation is that for databases that involve complex weighting procedures and statistical adjustments, only a small number of researchers may have the knowledge and skills to do the aggregation appropriately to permit reliable and valid estimates to be produced.

Strategy 1.d: Fund and conduct additional methodological studies to develop new approaches to using small samples for estimating AI/AN and NH/PI health and well-being.

Additional statistical methodologies, such as applying small area estimation techniques to small population groups, could be valuable and permit maximum use of available data at low cost. Small area estimation techniques are used when the sample size for the area or population group of interest within a larger study is too small to generate accurate estimates. These techniques involve the use of additional data (i.e., outside the targeted database) to supplement the data collected by the study to produce the estimates.

Strategy 1.e: Explore the potential for developing partnerships with AI/AN and NH/PI communities to conduct local area versions of major federal surveys that could supplement national survey data.

There are a number of examples of AI/AN tribes and communities that have initiated surveys of their populations, using sample designs, survey instruments, survey methodologies of major federal surveys [e.g., Navajo National Health Interview Survey (NHIS) and Youth Risk Behavior Study (YRBS)⁴⁶; Montana-Wyoming Tribal Leaders Council’s Consumer Assessment of Health Providers and Systems (CAHPS) and YRBS] and working in partnership with researchers and survey methodologists. These partnerships have been successful in replicating national surveys and have achieved response rates of 65 percent to 70 percent.⁴⁷ The development of Tribal Epidemiology Centers, funded by IHS, and a number of federal and private sector programs are increasing the number of emerging and new Native and other researchers resulting in greater capacity for data collection under the sponsorship of tribal organizations. With federal input and guidance and in partnership with universities or other research organizations, AI/AN and NH/PI communities could expand initiatives to collect data that could supplement national survey samples. Weighting of national survey data would mean that these supplemental data would likely not significantly affect national estimates; however, taking into account appropriate privacy and confidentiality rules, AI/AN and NH/PI communities that conduct these surveys could participate in and contribute to a national data repository that would be available to the AI/AN and NH/PI communities and to researchers for analysis and reporting. This strategy could be cost-effective, assuming the federal government would not pay for the data collection but would provide technical support and guidance to communities that would ensure that the data collected were comparable and consistent with national survey efforts.

Issue 2: Geographic Dispersion and Rural Concentration of the AI/AN and NH/PI Populations

Strategy 2.a: Develop and implement new methods for cost-effective identification of AI/AN and NH/PI who reside outside reservations/homelands to permit cost-effective sampling.

Geographic dispersion of the AI/AN and NH/PI population living off reservations and away from NH/PI homelands makes it difficult and costly to identify members of these population groups for sampling/over-sampling purposes. Strategies that might be developed and tested for the AI/AN and NH/PI populations include overlapping sample frames⁴⁸ and network

⁴⁶ The Navajo YRBS is a joint Navajo-BIA project.

⁴⁷ See, for example, Benally, C. et al., 2003 Navajo Middle and High School Youth Risk Behavior Surveillance System Report, Navajo Area Indian Health Service and the Navajo Nation (undated), which reports individual response rates of 67 percent to 70 percent. Similarly, the Montana-Wyoming Tribal Leaders Council’s Tribal Health Care Satisfaction Survey, conducted with adults on six reservations, has achieved response rates of 60 percent to 65 percent. Personal communication with the Executive Director of the MT-WY TLC.

⁴⁸ Overlapping sample frames is a technique that involves comparing separate lists of individuals to create a single frame that targets the population of interest. For example, researchers may compare a list of households to membership lists of tribes in order to identify households that are more likely to have AI/AN individuals. These identified households are then oversampled.

sampling⁴⁹ to assess their effectiveness and cost-effectiveness for these population groups. Another potential strategy might be drawing AI/AN and NH/PI samples from the respondents to existing surveys that identified race as AI/AN and NH/PI. For example, respondents to the National Health Interview Surveys who indicated race as AI/AN or NH/PI over several years could be identified by name and address to form a new sample frame for use in other DHHS surveys for which researchers wish to draw an over-sample of the AI/AN or NH/PI populations. Several researchers and interviewees suggested that the American Community Survey sample would be most useful for this purpose, due to its large sample size, but expressed concerns that confidentiality issues with Census-related surveys might prohibit this use.

Strategy 2.b: Increase rural sample sizes or over-sample AI/AN/NA in these areas to ensure that data on AI/AN residing on reservations and NH residing on homelands are available.

AI/AN populations are concentrated in rural areas (unlike the geographic dispersion observed in urban areas). Increasing rural sample sizes overall in federal surveys would increase AI/AN and NH/PI sample sizes, but would be costly to implement. A related strategy would be to over-sample these populations in areas where they are concentrated (e.g., reservations and NH homelands). As an example, if there was interest in over-sampling the AI/AN population for the Youth Risk Behavior Survey, states could choose to conduct surveys of AI/AN youth who reside on or near reservations or Alaska Native villages.⁵⁰ This approach also would involve increased costs, but would not be as costly as strategies to over-sample these populations on a national level. The limitation of this approach, however, is that it would improve data availability only for AI/AN and NH who reside on reservations or in homelands and would not address the under-representation in surveys of AI/AN and NH/PI who have migrated to other areas.

Strategy 2.c: Fund and conduct research to develop new methodological approaches that would make maximum use of existing data from rural areas, while protecting privacy and confidentiality of respondents.

Small sample sizes in rural areas raise the problem of protecting the privacy of the respondent and confidentiality of their responses. This issue reduces the availability of data on AI/AN and NH/PI residing in rural areas; thus, it would be useful to support research that could develop statistical or other methodological approaches that would permit greater use of data from these areas, within the requirements to protect privacy and confidentiality. If sample size permits and address information is available, administrative data could be geocoded and linked with socioeconomic variables available from the Census (at the block, tract, or county level). This method would help identify environmental factors that contribute to health and well-being disparities.⁵¹

⁴⁹ Network sampling uses the social network of respondents. For example, interviewers can ask an AI/AN respondent living outside the reservation whether they know of any other AI/AN individuals living outside the reservation, and then try and complete interviews with the new names obtained from the respondent.

⁵⁰ At least one state has implemented the YRBS in schools serving AI reservation children; see State of Montana Department of Health and Human Services website.

⁵¹ National Committee on Vital and Health Statistics, Subcommittee on Populations. Eliminating Health Disparities: Strengthening Data on Race, Ethnicity, and Primary Language in the United States. Hyattsville, MD: August 2005.

Issue 3: Misclassification of Race

Strategy 3.a: Develop new requirements and guidelines for accurate reporting of race/ethnicity on vital statistics records and administrative databases and provide training for those who are responsible for reporting.

The quality of the available data on AI/AN/NA health and well-being is affected by the accuracy with which race is reported on death certificates, registries, disease surveillance systems, administrative records, and other databases. There is evidence that there is substantial misclassification of AI/AN race on death certificates and that the rate of misclassification varies geographically. Improving data quality for vital statistics, registries, disease surveillance systems, and administrative records could be accomplished if requirements for race/ethnicity reporting were developed, guidelines disseminated, and training provided to those who are responsible for recording race into the databases. An example of one approach to improving race classification on death certificates and cancer registries would be to conduct matching of Indian Health Service files with state death certificate names or with cancer registry files to identify AI/AN people who have been racially misclassified. This approach would improve race identification on these databases, but only for AI/AN who are members of federally recognized tribes. A broader approach that would involve wide dissemination of guidelines and training of the wide range of people who are responsible for race identification on these databases would be more costly. However, while there would be an initial cost for this strategy, once the new requirements were developed and promulgated, training was provided, and quality reviews had been conducted to ensure that the new approach was successfully implemented, there would be minimal ongoing costs to improve these data.

Issue 4: Lack of/Inconsistent Collection of Race Identifiers in Some Data Sources

Strategy 4.a: Develop guidelines and encourage or require state agencies to collect uniform race/ethnicity data for all federally funded and federal-state jointly funded programs.

There is a lack of data on a range of health and well-being issues that are the focus of a number of federally funded or joint federal-state funded programs that are administered by the states. Some states collect race/ethnicity data, others do not; and reporting of race/ethnicity across states is inconsistent. Although OMB Directive 15 requires federal agencies to collect data on race and ethnicity in a consistent manner, this does not apply to the states. As a result, there are major gaps in availability and quality of data on AI/AN and NH/PI participation in and benefits from these programs. DHHS and other federal agencies could develop guidelines consistent with OMB Directive 15 and training manuals for race/ethnicity reporting. These agencies could also encourage or require states and local agencies that are responsible for implementing these programs to collect and report race and ethnicity in a consistent manner. The costs of this strategy would not be high, although reporting requirements could impose costs on

states and local agencies if they were to involve developing or modifying existing electronic reporting systems.

Issue 5: Inadequate Racial Representation, Limited Response Rates, and Question Interpretation

Strategy 5.a: Review sample designs for existing surveys to identify their potential to include AI/AN and NH/PI representatively and develop approaches that could increase representation.

Sample designs that rely on telephone numbers and addresses may disproportionately exclude AI/AN and NH/PI from selection for surveys. For example, the Behavioral Risk Factor Surveillance System (BRFSS) implements a telephone data collection approach. The methodological documentation for this study includes a discussion of the potential impact of varying rates of telephone coverage among different subgroups. Research on the magnitude of this sample design effect would be useful and if the magnitude is substantial, new approaches to supplement the regular sample with additional AI/AN and/or NH/PI could be developed and implemented. The cost of this strategy, for some specific surveys, however, could be significant.

Strategy 5.b: Fund and conduct additional research on approaches that could increase AI/AN and NH/PI response rates and test/implement these approaches.

Response rates to federal surveys for these populations appear to be lower than their probability of selection for the sample. While a number of factors have been identified as contributing to lower response rates (e.g., language/literacy levels, lack of access to telephones, high mobility), additional community-based research with AI/AN and NH/PI groups could be useful to identify new strategies for increasing responses. The increasing emphasis on Community-Based Participatory Research (CBPR) that involves partnerships between AI/AN communities and academic researchers is producing information on effective data collection strategies that produce high response rates.⁵² A systematic review of the approaches used in CBPR, successes, and their costs could provide useful information on strategies that might be adopted more widely by federal surveys. These could then be tested for effectiveness (and cost-effectiveness) and implemented to increase data availability and quality.

Strategy 5.c: Conduct ongoing cognitive testing of current and new survey instruments to assess the extent of cultural differences in question interpretation and develop alternative wording, if necessary.

The quality of the data available on AI/AN/NA health and well-being is affected, if members of these population groups interpret survey questions differently than other

⁵² One example of CBPR involving partnerships between AI/AN communities and academic researchers discussed elsewhere in this report is the Tribal Health Care Satisfaction Survey currently being conducted by the Tribal Leaders Council of Montana and Wyoming and Black Hills State University under an AHRQ Minority Research Infrastructure Support Program grant. This survey, conducted with adults on six reservations, has achieved response rates of 60 percent to 65 percent.

racial/ethnic group members. While cognitive testing is conducted routinely for many ongoing federal surveys, focusing these efforts on the AI/AN and NH/PI populations would identify the extent to which cultural differences are affecting question interpretation and permit changes to wording that would improve the quality of available data. However, there are over 560 federally recognized tribes and a number of NH/PI subgroups, each of which may differ culturally. The costs of conducting cognitive testing with each subgroup would likely be prohibitive, as would modification of questionnaires for each group. Cognitive testing to assess commonalities across all AI/AN groups might produce useful information for national surveys and cognitive testing of specific population subgroups to support regional, state, or other targeted population surveys could be useful, nevertheless.

Issue 6: Exclusion of Pacific Insular Areas from Some Surveys

Strategy 6.a: Include Pacific Insular Areas in most relevant surveys and other data collection, either routinely or periodically.

The geographic distance of the Pacific Insular Areas (those that are outside of Hawaii) imposes additional costs for surveys and other data collection efforts, if they are included. These areas could be included in the most relevant surveys—that is, those surveys that would fill major gaps in health and well-being data—either on a regular basis or periodically, in order to improve the availability of data for Pacific Islanders residing in these areas. The limitation of this strategy, obviously, is that it would involve additional costs. These costs could be minimized by conducting these surveys on a periodic basis, rather than every year.

Issue 7: Inadequate Collection of Data on AI/AN/NA Subgroups

Strategy 7.a: Refine racial identifiers to better focus on subgroups of AI/AN and NH/PI populations, and design and implement data collections to obtain sample sizes sufficient for analysis of health and well-being for specific subgroups.

There are very few data sets that have detailed identifiers of AI/AN tribal affiliation or NH/PI subgroups. Strategies for improving availability and quality of data for subgroups of the AI/AN/NA population would necessarily be similar to those discussed above, but would require including more detailed racial/subgroup identifying questions (and guidance for reporting race on vital statistics records, disease registries, etc.), considerably greater over-sampling of the population, creative strategies for identifying potential sample members, or targeted new data collection. These strategies would likely be very costly to implement and funding may be difficult to obtain, without reducing survey efforts in other areas. Other possible strategies for improving data availability for these subgroups that could be considered include the community-level data collection approach discussed above under the Small Population Size issue. To obtain data pertaining to members of federally recognized tribes, the Indian Health Service and Bureau of Indian Affairs data collection and reporting capacities could be strengthened.

Discussion

Most of these strategies have been identified in previous studies of AI/AN and NH/PI data availability and quality issues. In addition, a number of them have been initiated by DHHS and other federal agencies, either in the past or as current initiatives. These are discussed in the next section. It is important, also, to recognize that the individual strategies discussed above may result in only marginal improvements in AI/AN/NA data availability and quality. Major improvements would likely require that multiple strategies be implemented concurrently or over time in order to fully address the wide range of issues that currently affect AI/AN/NA data availability.

5.3 Current and Planned Initiatives of DHHS and Other Federal Agencies that Would Improve Availability and Quality of Data on AI/AN/NA Health and Well-being

Over the past two decades, DHHS and its operating divisions have given considerable attention to identifying issues and developing initiatives to improve the availability and quality of data for measuring, monitoring, and research on racial and ethnic health disparities. Much of the selected literature and reports that were reviewed in the earlier section of this report have been funded by DHHS or represent internal DHHS efforts to develop strategies for improving AI/AN/NA health and well-being data availability and quality (as well as data for other racial/ethnic groups). In this section, current and planned initiatives of DHHS and other agencies that are designed to improve data for AI/AN/NA population groups are described and categorized within the set of strategies discussed above. Most of the information in the section was obtained through interviews with DHHS staff and other knowledgeable individuals. Because only a small number of individuals were interviewed, this list of current and planned initiatives is not comprehensive and it is likely that there may be other major initiatives that were not identified.

Issue 1: Small Population Size

Strategy 1.a: Increase sample sizes of federal surveys to ensure that sufficient samples are obtained to permit analysis of AI/AN and NH/PI health and well-being issues.

No known current or planned initiatives were discussed in the selected interviews. Although several interviewees mentioned increasing sample sizes to obtain larger numbers of AI/AN and NH/PI respondents, they cited the costs of this approach as a major reason that it would not be feasible.

Strategy 1.b: Over-sample the AI/AN and NH/PI population in federal surveys in order to obtain sufficient sample sizes for these groups.

The National Health Interview Survey is implementing over-samples of Blacks and Hispanics to improve data availability and quality; however, there are no immediate plans for an over-sample of the AI/AN and NH/PI populations, either on a one-time or rotating basis. The Survey of American Indians and Alaska Natives (SAIAN), funded jointly by the Agency for Health Care Research and Quality (AHRQ) (then the National Center for Health Services Research) and IHS, was conducted in the late 1980s. The experience and lessons learned from the SAIAN provide useful background and information for consideration of new initiatives that involve over-sampling of the AI/AN population, particularly for those initiatives that might focus on increasing data availability for members of federally recognized tribes. For example, SAIAN required consultation with tribes, hiring and training of native language speakers to conduct interviews, and extensive travel to the reservations and Alaska Native villages included in the survey. These efforts extended the time necessary for data collection and required higher costs per completed interview than are typical for most surveys. Discussions have been underway between the AHRQ Medical Expenditure Panel Survey staff and IHS on the possibility of conducting another SAIAN. However, funding may not be available to permit this in the near future.

Strategy 1.c: Aggregation of multiple years of data from surveys that are conducted annually to obtain sufficient AI/AN and NH/PI sample sizes.

Several operating divisions within DHHS are exploring the potential for aggregation of data to increase the potential for examining AI/AN and NH/PI health and well-being. An example of the application of this approach is described in the 2005 paper *Health Characteristics of the American Indian and Alaska Native Adult Population: United States, 1999-2003*,⁵³ that aggregates five years of NHIS data to analyze AI/AN health. Another example is the AHRQ Medical Expenditure Panel Survey analysis of aggregated data to produce a chartbook on racial and ethnic differences in health insurance coverage and usual sources of care.⁵⁴ Finally, the Census Bureau's American Community Survey is structured so that the cumulative data over several years will provide information on small areas such as counties, reservations, and the Native Hawaiian homelands.

Strategy 1.d: Fund and conduct additional methodological studies to develop new approaches to using small samples for estimating AI/AN and NH/PI health and well-being.

⁵³ Barnes, P.M., Adams, P.F. and Powell-Griner, E., "Health Characteristics of the American Indian and Alaska Native Adult Population: United States, 1999-2003," *Advance Data from Vital and Health Statistics*, Number 356, April 27, 2005.

⁵⁴ Roberts, M. "Racial and Ethnic Differences in Health Insurance Coverage and Usual Source of Health Care, 2002." Rockville (MD): Agency for Healthcare Research and Quality; 2006. MEPS Chartbook No. 14. AHRQ Pub. No. 06-0004.

DHHS has funded a number of methodological studies to develop new strategies for using existing data more effectively, including:

- Assessment of Major Federal Data Sets for Analyses of Hispanic and Asian or Pacific Islander Subgroups and Native Americans: Task 3 Report: Extending the Utility of Federal Databases by J. Waksberg, D. Levine, and D. Marker (prepared for DHHS/ASPE: Westat, Rockville, MD, May 2000);
- Rural Research Needs and Data Sources for Selected Human Services Topics by D.A. Strong, et al. (prepared for U.S. DHHS/ASPE by Mathematica Policy Research, Princeton, NJ, August 2005); and
- Draft Report: Evaluation of Statistical Methods for Data Collection and Analysis on Racial and Ethnic Minorities and Other Hard-to-Reach Populations by B.K. Finch, et al. (prepared for U.S. DHHS/OMH). This report can be located at: <http://paa2006.princeton.edu/download.aspx?submissionId=61646>.

In addition, internal DHHS staff are conducting research on these methodological issues. Examples include:

- The National Center for Health Statistics (NCHS) has developed a “bridging” algorithm for the National Health Interview Survey and other data analysis that assigns race based on probabilities. This permits comparison of new multi-race reporting with old single-race reporting.
- AHRQ is currently working to augment Healthcare Cost and Utilization Project (HCUP) hospital data with IHS hospital data to get a better picture of hospitalization rates for AI/AN. In current HCUP hospital data, the estimated rates for AI/AN are lower than expected. When these two data sources are combined, the numbers are more in line with hospitalization rates for the general population.

Strategy 1.e: Explore the potential for developing partnerships with AI/AN and NH/PI communities to conduct local area versions of major federal surveys that could supplement national survey data.

Several initiatives to conduct local area versions of major federal surveys have been funded by DHHS, primarily through grants. Examples include development and implementation of the Consumer Assessment of Health Plans and Providers (CAHPS) with tribes in Montana and Wyoming under an AHRQ Minority Research Infrastructure Support Program grant, and implementation of the Youth Risk Behavior Survey with tribes in Montana under a National Institutes of Health (NIH)/National Center for Minority Health and Health Disparities (NCMHD)-funded Center of Excellence in Partnerships for Community Outreach, Research on Health Disparities, and Training (Project EXPORT) grant. In addition, the Navajo Tribe has replicated the National Health and Nutrition Survey with assistance from NCHS. Several of the regionally based Tribal Epidemiology Centers funded by the Indian Health Service include projects to improve data on the health of AI/AN populations within their IHS area using federal

survey questionnaires.⁵⁵ In addition, there are a number of programs funded by federal agencies that provide support for development of AI/AN and NH/PI researchers.⁵⁶

Issue 2: Geographic Dispersion and Rural Concentration of the AI/AN and NH/PI Populations

Strategy 2.a: Develop and implement new methods for cost-effective identification of AI/AN and NH/PI who reside outside reservations/homelands to permit cost-effective sampling.

The California Health Interview Survey identifies urban areas with significant concentrations of AI/AN (e.g., areas around urban Indian health centers) and draws samples from these areas to increase AI/AN representation in the surveys.

Strategy 2.b: Increase rural sample sizes or over-sample AI/AN in these areas to ensure that data on AI/AN residing on reservations and NH/PI residing on homelands are available.

No known current or planned initiatives were discussed during the selected interviews, other than those described above under Issue: Small Population Size: Strategy 2.

Strategy 2.c: Fund and conduct research to develop new methodological approaches that would make maximum use of existing data from rural areas, while protecting privacy and confidentiality of respondents.

No known current or planned initiatives were discussed during the selected interviews, other than those described above under Issue: Small Population Size: Strategy 4.

Issue 3: Misclassification of Race

Strategy 3.a: Develop new requirements and guidelines for accurate reporting of race/ethnicity on vital statistics records and administrative databases and provide training for those who are responsible for reporting.

IHS and the Centers for Medicare and Medicaid Services have collaborated to match IHS patient users with CMS beneficiaries to improve racial identification. One IHS Area Office is known to work with states to conduct matches of IHS patients with state cancer registries to improve racial identification. IHS is doing a second study on racial misclassification of mortality data and developing adjustments to apply to the National Death Index data to compensate for the misreporting of AI/AN race on state death certificates. Vital statistics will be reported with multiple races, using categories established by OMB Directive 15. This is being phased in, and

⁵⁵ The Indian Health Service divides the country into 12 areas corresponding to regions of the country for provision of health care services.

⁵⁶ Examples are the NIH/NCMHD Project EXPORT grants, the IHS-NIH Native American Research Centers on Health grants, the AHRQ Minority Research Infrastructure Support Program grants, and the National Institute on Aging Resource Centers on Minority Aging Research, all of which include components to encourage and support new minority researchers.

five states currently report multiple races for the mother on birth certificate. Over the next few years, between 15 and 20 states will have multi-race options on their birth certificates. The phase-in is necessary because states will need to revise their electronic submission systems to report multiple races. It is anticipated that all states will submit multiple race data on birth certificates by the end of the decade.

Issue 4: Lack of/Inconsistent Collection of Race Identifiers in Some Data Sources

Strategy 4.a: Develop guidelines and encourage or require state agencies to collect uniform race/ethnicity data for all federally funded and federal-state jointly funded programs.

No known current initiatives were discussed during the selected interviews other than for vital statistics records (see above). Due to OMB Directive 15, it is possible that the vital statistics approach will expand to joint federal-state programs over time.

Issue 5: Inadequate Racial Representation, Limited Response Rates, and Question Interpretation

Strategy 5.a: Review sample designs for existing surveys to identify their potential to include AI/AN and NH/PI representatively and develop approaches that could increase representation.

Department of the Interior (DOI) is working with Census to consider the feasibility of including the PI population residing in Pacific Insular Areas in the American Community Survey.

Strategy 5.b: Fund and conduct additional research on approaches that could increase AI/AN and NH/PI response rates and test/implement these approaches.

Community-based participatory research projects, involving academic-AI/AN/NA community partnerships, are producing information on approaches that are effective for increasing response rates to surveys. Many of these projects are grant-funded through DHHS agencies and the information generated could provide valuable insights for other federal surveys. Census Bureau conducts research on strategies for increasing AI/AN/NA response to mail surveys.

Strategy 5.c: Conduct ongoing cognitive testing of current and new survey instruments to assess the extent of cultural differences in question interpretation and develop alternative wording, if necessary.

NCHS is conducting cognitive testing to investigate whether survey instruments used in national surveys are appropriate for subpopulations. Their findings suggest that specific questions (e.g., health status questions) may be interpreted differently by different populations, including AI/AN individuals.

Issue 6: Exclusion of Pacific Insular Areas from Some Surveys

Strategy 6.a: Include Pacific Insular Areas in most relevant surveys and other data collections, either routinely or periodically.

The Department of the Interior (DOI), Office of Insular Affairs has a current initiative to improve data for residents of the Pacific Insular Areas, both those residing in U.S. territories and those residing in the Freely Associated States. DOI is funding the National Center for Health Statistics through a Memorandum of Agreement between DOI and DHHS to conduct a study, Inventory and Assessment of Health Information Sources in the U.S. Insular Areas, of available data, gaps in the data, and strategies that can be utilized for data improvement.

Issue 7: Inadequate Collection of Data on AI/AN/NA Subgroups

Strategy 7.a: Refine racial identifiers to better focus on subgroups of AI/AN and NH/PI populations, and design and implement data collections to obtain sample sizes sufficient for analysis of health and well-being for specific subgroups.

The Census Bureau, in its preparation for the 2010 Census, is testing the feasibility of improving tribal reporting, i.e., asking each respondent to identify principal or enrolled tribe. Census Bureau is also conducting focus groups to better understand how to present tribal data, e.g., whether to present data separately for federally recognized tribal members, state-recognized tribes, and/or bands or clans within larger tribal groups. In addition, the California Health Interview Survey collects data on respondents' specific tribal affiliations, but sample sizes are too small for analytic purposes without multi-year aggregation. The Indian Health Service is developing a data warehouse that will include individual records of IHS patients and will permit queries; this may increase the availability of health data for federally recognized tribal members. The IHS also is developing a behavioral health database that will provide improved data on behavioral health issues for this population.

6 SUMMARY AND CONCLUSIONS

6.1 Summary of Findings

The analysis of data availability and quality for measuring, monitoring, and conducting research on AI/AN/NA health and well-being found that there are gaps in available data and a paucity of data sets to address some well-being issues, for the combined AI/AN and NH/PI population groups. Gaps in the data available to examine health and well-being for subgroups of the AI/AN/NA population to address health and well-being issues are greater. Major findings of this analysis of data sets with sample sizes of at least 200 include:

- For the combined AI/AN population, there are more than two data sets for each policy area, with the exception of military/veterans issues.
- For the combined NH/PI population, five of the ten policy areas are covered by data sets; there are limited data (i.e., two or fewer data sets) for education, elder well-being, justice system issues, and transportation issues; and no data sets are available for examination of military/veterans issues.
- For the AI and AN populations separately, there are no data sets available to examine child well-being, elder well-being, justice system issues, and military/veterans issues and limited data available for each of the other policy areas.
- For the NH group separately, there are more than two data sets available to examine health issues, limited data available to examine economic well-being, education, family well-being, housing, justice system issues, and transportation issues; no data sets are available to examine child-well-being, elder well-being, or military/veterans issues.
- For the PI group separately, there are gaps in data availability for child-well-being, elder well-being, justice system issues, and military/veterans issues and limited data available for examination of each of the six other policy areas.

While we found no gaps in the available data for the combined AI/AN group and only one gap for the combined NH/PI data, there are a substantial number of policy areas for which data are not available to examine health and well-being for the separate AI, AN, NH, and PI groups. It is important, also, to remember that the data sets examined may vary in the indicators available to examine specific issues within a policy area. For some policy areas and issues, there may be inadequate data for analysis of disparities in health and well-being for these population groups, by demographic, geographic, or economic characteristics, even though some data on the issue/policy area may be collected and usable for aggregate analysis. For example, some data sets focus on specific geographic regions (e.g., California or Hawaii), while others can only be used to produce national-level estimates. There may be substantial differences in health and well-being by region of the country, urban and rural location, and other characteristics of AI/AN and NH/PI subgroups of the population. There are very few databases that permit these geographic and subgroup differences to be measured, although such measures may be important for accurately assessing disparities and for identifying strategies to improve health and well-being.

A number of possible strategies for improving the availability and quality of data on AI/AN/NA health and well-being have been presented and discussed in this report, drawn from a review of selected literature and interviews with a limited number of knowledgeable individuals. Many of these strategies have been initiated within DHHS, either partially or fully. Because of the limitations of this study, we have less information on current or planned initiatives within other federal agencies, but it is likely that there are additional efforts underway that are not reflected in this report.

In Table 6, at the end of this section, we summarize the issues, strategies, possible levels of costs for implementation, and current/planned initiatives to improve AI/AN/NA data that have been identified in this study. The potential costs of each strategy are ranked as “Low,”

“Medium,” or “High” to reflect the additional funds that might be necessary to implement the strategy. These additional costs could vary substantially depending on specific surveys and databases that might be affected and sample sizes that would be involved. Although these potential costs are relative and not meant to be precise measures of each strategy’s cost, it would also be important to weigh the potential costs of strategies against the benefits of improved data collection.

Only three of the strategies were identified as having no current or planned initiatives underway that would, at least partially, address the associated issue. These are: 1) increasing sample sizes of relevant federal surveys; 2) over-sampling of AI/AN/NA within existing relevant federal surveys; and 3) increasing sample sizes for AI/AN/NA residing on reservations, AN villages, and NH homelands. Each of these strategies is also identified as potentially involving high additional costs to implement. In addition, the strategy to improve PI data by including the Pacific Insular Areas in relevant federal surveys is under study, but is not actually being implemented as a strategy. Again, the cost of implementation of this strategy is likely to require medium to high additional costs, depending on whether these areas were included periodically or routinely in federal surveys. Current or planned initiatives are underway for all of the other strategies, either within federal agencies, states, or through academic-native community partnerships. None of these strategies is ranked as requiring high additional costs and most are ranked as of low or low/medium costs.

Although there are a substantial number of initiatives underway or planned in the future, it is important to recognize that few of the issues that are important to improving data availability and quality are fully addressed by the current initiatives. For instance, strategies which focus on research to identify new statistical approaches for small area/population estimation may produce potential methodologies that could improve the usefulness of the available data, but implementation of those methodologies might involve complex and potentially costly software development and analytic time that would be prohibitive. Similarly, research and cognitive testing to identify differences among populations in understanding and interpretation of survey questions is a relatively low-cost strategy; however, implementation of strategies to address these differences and improve data quality may involve high additional costs.

The methodology used for this study has some limitations. Although only 67 data sets comprise the foundation for this report, it is unlikely that there are many other federal or other publicly available databases that have sufficient sample sizes of AI/AN/NA and include measures of health and well-being that would substantially affect the findings of this analysis. Because it was not possible for us to fully examine the quality of individual databases, it is possible that analysts and researchers who use these data sets may identify quality issues that will result in additional gaps pertaining to the health and well-being policy areas. Finally, it is very likely that the limited number of knowledgeable people who were interviewed were not aware of all of the current and planned initiatives within DHHS and other organizations. Thus, there likely are more activities underway and planned than are presented in this report that could contribute to improved data on AI/AN/NA health and well-being. Given these caveats, this report provides the reader an overview of gaps in AI/AN/NA data, strategies to improve data, and some information on current initiatives that may be valuable in leading to improved data in the future.

6.2 Conclusions

A comprehensive approach to addressing the data gaps identified could involve the following steps:

- A DHHS-wide coordinated approach to implement many of the identified strategies across DHHS agencies that both use and collect data on health and well-being issues. Such a coordinated approach would involve soliciting information from these agencies on the key issues for measuring, monitoring, and research on health and well-being disparities and issues, sharing of information gained from current and planned initiatives across DHHS agencies, and a process for determining specific approaches that would be adopted and used consistently in DHHS surveys and administrative databases. This could be accomplished through an incremental approach that begins with coordination among several key DHHS agencies, with the DHHS Data Council perhaps taking a lead role in the process. This effort could then be expanded over time to involve additional DHHS agencies in the process.
- Coordination and sharing of results of current initiatives among all federal agencies that use and collect data on AI/AN/NA health and well-being. This coordination and sharing of finding would involve development of a process for adopting and implementing consistent data collection strategies to improve these data.
- Working with states to improve the completeness, accuracy, and consistency of collection of race identifiers to ensure that vital statistics and administrative databases contain accurate identification of AI/AN/NA people.
- Consultation and involvement of AI/AN/NA tribes and communities as partners in the process of improving these data, to ensure that strategies that are identified and implemented have the support of these populations and that the approaches that will be implemented also reflect the priorities of the population that will benefit from improved data.

If the coordination strategies outlined above are implemented, it is likely that the availability and quality of data on AI/AN/NA health and well-being will substantially improve. This outcome will require that a long-run focus be maintained on the need for and importance of data to measure, monitor, and analyze disparities in health and well-being of this population. In addition, consideration of the potential benefits relative to costs of alternative strategies may be useful for guiding the decisions that will support these important data improvements.

Table 6. Summary of Strategies, Potential Costs, and Current/Planned Initiatives

Issues and Strategies	Summary of Strategy	Potential Cost Level of Strategy (Low, Medium, High)	Current/Planned Initiative
<i>Issue 1: Small Population Size</i>			
Strategy 1.a	Increase sample sizes for relevant federal surveys	High—would require substantial increase in sample size for most federal surveys	No known current or planned initiative.
Strategy 1.b	Over-sample AI/AN/NA	High—for nationally representative surveys Medium or Low—for surveys focused on reservations or NH homelands	No known current or planned initiative.
Strategy 1.c	Aggregate multiple years of data	Low	National Health Interview Survey (NHIS), MEPS, CAHPS have all aggregated data to examine AI/AN/NA populations. American Community Survey will use multiple years of data collection for analysis of small populations.
Strategy 1.d	Conduct research on statistical methods for using small samples	Low/Medium	ASPE and OMH have funded methodological studies. NCHS and AHRQ have conducted methodological research on small sample populations.
Strategy 1.e	Federal-community partnerships to replicate federal surveys	Low/Medium—depending on source of funding for implementing surveys	AHRQ and NIH have funded grants that include replication of federal surveys with AI/AN populations. NCHS has provided technical assistance to tribal NHIS. IHS-funded Tribal Epidemiology Centers are replicating some surveys.
<i>Issue 2: Geographic Dispersion/Rural Concentration</i>			
Strategy 2.a	Develop methods for low-cost identification of members of dispersed populations	Low—particularly if focus on developing agreements to draw samples from prior survey respondents who have identified themselves as AI/AN/NA	California HIS has tested methods for identifying urban areas with AI/AN concentrations.
Strategy 2.b	Increase sample sizes of AI/AN/NA residing on reservations and NH homelands	Medium/High—depending on sample size and survey methodology	No known current or planned initiatives.

**Table 6. Summary of Strategies, Potential Costs, and Current/Planned Initiatives
(Continued)**

Issues and Strategies	Summary of Strategy	Potential Cost Level of Strategy (Low, Medium, High)	Current/Planned Initiative
<i>Issue 3: Misclassification of Race</i>			
Strategy 3.a	Develop requirements and guidelines for improving race identification on state vital statistics and administrative databases	Low—for federal government Medium—initial costs to states could be substantial if new data systems or substantial modification of existing systems are required	IHS and federal and state agencies conduct record matches. States are adopting OMB Directive 15 with assistance from NCVHS to improve race identification on birth records. IHS is conducting a study to identify the extent of misclassification of AI/AN race on death records and variations in misclassification across states.
<i>Issue 4: Inconsistent Collection of Race Data</i>			
Strategy 4.a	Require states to collect uniform race data for federally funded programs	Low—for federal government Medium—costs to states	No known current or planned initiatives, other than vital statistics records.
<i>Issue 5: Racial Differences in Sample Selection, Response, and Question Interpretation</i>			
Strategy 5.a	Review sample designs of federal surveys and ensure/increase AI/AN/NA representation	Medium to High	No known current or planned initiatives.
Strategy 5.b	Research and test new strategies to increase AI/AN/NA response rates	Low—if research and testing relies on past or on-going participatory research Medium/High—if other types of new strategies are identified	Several DHHS grant programs are funding academic/community partnerships that are conducting surveys with new methods to increase response rates.
Strategy 5.c	Conduct cognitive testing to assess extent to which cultural differences affect interpretation of questions	Low—much cognitive testing is already conducted for many ongoing federal surveys	NCHS and other DHHS survey groups, as well as Census, conduct cognitive testing.

**Table 6. Summary of Strategies, Potential Costs, and Current/Planned Initiatives
(Continued)**

Issues and Strategies	Summary of Strategy	Potential Cost Level of Strategy (Low, Medium, High)	Current/Planned Initiative
<i>Issue 6: Exclusion of Pacific Insular Areas from Surveys</i>			
Strategy 6.a	Include Pacific Insular Areas in relevant surveys, routinely or periodically	Medium/High—depending on sample sizes and frequency of data collection	DoI initiative is focused on improving data for residents of Pacific Insular Areas. As a first step, an Inventory and Assessment of Health Information Sources in the U.S. Insular Areas is being conducted.
<i>Issue 7: Inadequate Collection of Data on AI/AN/NA Subgroups</i>			
Strategy 7.a	Refine racial identifiers to include specific AI/AN/NA subgroups	Low	<p>Census Bureau is examining further refinement of AI/AN/NA subgroup identification and useful ways to present subgroup data. California HIS collects tribal-specific affiliation data.</p> <p>IHS is improving specific data sets, including behavioral health data, that will increase the availability and quality of data on the health of AI/AN who are members of federally recognized tribes.</p>

APPENDICES

Appendix A – List of Policy Areas and Issues

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, health 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, marital 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

1. Transportation quality (e.g., maintenance of roads, availability of paved roads, child restraint laws)
2. Transportation availability (e.g., availability of public transit, vehicle ownership per household)

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)

**Appendix B – Sample Sizes for Data Sets by Policy
Area, for Each Population Group**

Table B-1. Sample Sizes for the AI/AN Race Category⁵⁷

Child Well-being	Number of AI/AN	Field Period
Adoption and Foster Care Analysis and Reporting System (AFCARS)	10,260 foster care 2,190 waiting for adoption 700 adopted	2003
National Child Abuse and Neglect Data System (NCANDS)	46,708 children 6,294 perpetrators	2004
National Survey of Child and Adolescent Well-being (NSCAW)	341 CPS sample, 47 LTFC sample	1999- 2001
Runaway and Homeless Youth Management Information System (RHYMIS)	1,922	2004
Demographic and Economic Indicators	Number of AI/AN	Field Period
American Community Survey (ACS)	unknown	
Census 2000	2,475,956	2000
Census 2000 - The American Indian and Alaska Native Summary File	unknown	
Economic Well-being	Number of AI/AN	Field Period
Census of Agriculture	5,268	2002
Consumer Expenditure Surveys (CE) Interview and Diary Surveys	204 interview, 80 diary	2004
Current Population Survey (CPS)	1,486	April 2006
Panel Study of Income Dynamics (PSID)	136 men, 39 women	2003
Survey of Program Dynamics	1,100	1992- 2002
Education	Number of AI/AN	Field Period
Early Childhood Longitudinal Study, Birth Cohort (ECLS-B)	750	2001- 2002
Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K)	210 child assessments	2003- 2004
Head Start Program Information Report	55,733 (Program level data only)	2004- 2005
Integrated Postsecondary Education Data System (IPEDS)	170,919 (Program level data only)	2004
National Assessment of Adult Literacy (NAAL)	167	2003
National Household Education Surveys Program (NHES)	108 PFI-NHES 233 ECPP-NHES 355 AE-NHES 374 ASPA-NHES 193 Parent-NHES	2003 2005 2005 2005 1999

⁵⁷ For some data sets, exact counts of the number AI/AN/NA in the data are not available. However, for these data sets, review of the documentation indicated that there are at least 100 (usually more) AI/AN/NA represented in the data. This is true for all subgroups in the tables in this appendix (i.e., AI/AN, AI, AN, NH/PI, NH, and PI).

Table B-1. Sample Sizes for the AI/AN Race Category (continued)

National Indian Education Study (NIES)	3,800 grade 4, reading 3,400 grade 8, reading 3,900 grade 4, math 3,500 grade 8, math	2005
Elder Well-being	Number of AI/AN	Field Period
Health and Retirement Study (HRS)	162	1992
National Aging Program Information Systems (NAPI) State Performance Reports	56,606 (Program level data only)	2004
Family Well-being	Number of AI/AN	Field Period
Food Stamp Quality Control Database (FSPQC)	4,050	2004
National Survey of America's Families (NSAF)	491 focal children 1,018 paired adults 752 random adults 248 childless adults	2002
National Survey of Family Growth (NSFG)	368 non-Hispanic 579 Hispanic	2002
Temporary Assistance for Needy Families (TANF) and Tribal TANF	9,718 active 3,001 closed 9,983 tribal (2002)	2004
Health Policy Issues	Number of AI/AN	Field Period
Behavioral Risk Factor Surveillance System (BRFSS)	6,904	2005
CAHPS Health Plan Survey Response Data	1,745 Adult Commercial 28 Child Commercial 1,011 Adult Medicaid 1,184 Child Medicaid 5 SCHIP	2005
California Health Interview Survey (CHIS)	<u>Adults</u> 740 Hispanic adults 1,157 Non-Hispanic <u>Adolescents</u> 212 Hispanic 153 Non-Hispanic <u>Children</u> 195 Hispanic 175 Non-Hispanic	2003
Health Behavior in School-aged Children (HBSC)	572	2001-2002
Health Information National Trends Survey (HINTS)	141	2005
Medicaid Analytic Extract (MAX)	806,211	2003
Medical Expenditure Panel Survey (MEPS)	293	2004
Medicare Denominator Files	141,000	2002
Medicare Utilization–Standard Analytic Files (SAFs)	N/A	
National Ambulatory Medical Care Survey (NAMCS)	93	2004
National Epidemiologic Survey on Alcohol and Related Conditions (NESARC)	1,304	2001-2002
National Health Interview Survey (NHIS)	670	2005

Table B-1. Sample Sizes for the AI/AN Race Category (continued)

National Hospital Ambulatory Medical Care Survey (NHAMCS)	209 ER Dept. 149 Outpatient Dept.	2004
National Longitudinal Mortality Study (NLMS)	19,779	1979-1998
National Mortality Followback Survey (NMFS)	205	1993
Health Policy Issues (continued)	Number of AI/AN	Field Period
National Survey on Drug Use and Health (NSDUH)	784	2004
National Vital Statistics System: Linked Birth-Infant Death (NVSS-I)	43,054 births	2003
National Vital Statistics System: Mortality (NVSS-M)	13,160	2003
National Vital Statistics System: Natality (NVSS-N)	43,927	2004
Pediatric Nutrition Surveillance System (PedNSS)	77,915	2004
Pregnancy Nutrition Surveillance System (PNSS)	11,686	2004
Pregnancy Risk Assessment Monitoring System (PRAMS)	N/A	
Resource and Patient Management System (RPMS) and National Patient Information Reporting System (NPIRS)	N/A	
Surveillance, Epidemiology, and End Results (SEER)	28,000 cases involving AI/AN	cumulative
Tobacco Use Supplement to the Current Population Survey (TUS-CPS)	931	November 2003
United States Renal Data System (USRDS)	1,097	2003
Washington State Population Survey (WSPS)	265 alone 225 combined	2004
Youth Risk Behavior Surveillance Survey (YRBSS)	147	2005
Housing Issues	Number of AI/AN	Field Period
American Housing Survey (AHS)	300	2003
American Housing Survey: Metropolitan Surveys	350	2004
Justice System Issues	Number of AI/AN	Field Period
Annual Survey of Jails (ASJ)	N/A	
Census of Jails	N/A	
National Crime Victimization Survey (NCVS)	1,621	1992-2004
Uniform Crime Reports	135,479 arrests	2004
Military Service/Veterans Issues	Number of AI/AN	Field Period
National Survey of Veterans (NSV)	897	2001
Transportation	Number of AI/AN	Field Period
National Household Travel Survey (NHTS)	401 Head of HH 882 Person in HH 697 Primary drivers 3,383 Day travelers 277 Long-trip travelers	2001

Table B-2. Sample Sizes for the AI Alone Race Category

Demographic and Economic Indicators	Number of AI alone	Field Period
American Community Survey (ACS)	N/A	
Elder Well-being	Number of AI alone	Field Period
Health and Retirement Study (HRS)	N/A	1992
Health Policy Issues	Number of AI alone	Field Period
Resource and Patient Management System (RPMS) and National Patient Information Reporting System (NPIRS)	N/A	
Treatment Episode Data Set (TEDS)	38,785	2004

Table B-3. Sample Sizes for the AN Alone Race Category

Demographic and Economic Indicators	Number of AN alone	Field Period
American Community Survey (ACS)	N/A	
Elder Well-being	Number of AN alone	Field Period
Health and Retirement Study (HRS)	N/A	1992
Health Policy Issues	Number of AN alone	Field Period
Resource and Patient Management System (RPMS) and National Patient Information Reporting System (NPIRS)	N/A	
Treatment Episode Data Set (TEDS)	5,186	2004

Table B-4. Sample Sizes for the NH/PI Race Category

Child Well-being	Number of NH/PI	Field Period
Adoption and Foster Care Analysis and Reporting System (AFCARS)	1,540 foster care 340 waiting for adoption 130 adopted	2003
National Child Abuse and Neglect Data System (NCANDS)	11,700 children 2,091 perpetrators	2004
Runaway and Homeless Youth Management Information System (RHYMIS)	338	2004
Demographic and Economic Indicators	Number of NH/PI	Field Period
American Community Survey (ACS)	unknown	
Economic Well-being	Number of NH/PI	Field Period
Census of Agriculture	280	2002
Consumer Expenditure Surveys (CE) Interview and Diary Surveys	118 Interview 55 Diary	2004
Current Population Survey (CPS)	486	April 2006

Table B-4. Sample Sizes for the NH/PI Race Category (continued)

Education	Number of NH/PI	Field Period
Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K)	144 child assessments	2003-2004
Head Start Program Information Report	8,448 (Program level data only)	2004-2005
National Assessment of Adult Literacy (NAAL)	26	2003
National Household Education Surveys Program (NHES)	66 ECPP-NHES 51 AE-NHES 79 ASPA-NHES	2005
Family Well-being	Number of NH/PI	Field Period
Food Stamp Quality Control Database (FSPQC)	N/A	
National Survey of Family Growth (NSFG)	91	2002
Temporary Assistance for Needy Families (TANF) and Tribal TANF	1,711 active 653 closed	2004
Health Policy Issues	Number of NH/PI	Field Period
Behavioral Risk Factor Surveillance System (BRFSS)	1,503	2005
CAHPS Health Plan Survey Response Data	1,006 Adult Commercial 9 Child Commercial 194 Adult Medicaid 1,354 Child Medicaid 4 SCHIP	2005
California Health Interview Survey (CHIS)	<u>Adults</u> 61 Hispanic 199 Non-Hispanic <u>Adolescents</u> 27 Hispanic 39 Non-Hispanic <u>Children</u> 34 Hispanic 57 Non-Hispanic	2003
Health Behavior in School-aged Children (HBSC)	116	2001-2002
Health Information National Trends Survey (HINTS)	17	2005
Medicaid Analytic Extract (MAX)	508,106	2003
Medical Expenditure Panel Survey (MEPS)	150	2004
National Ambulatory Medical Care Survey (NAMCS)	70	2004
National Epidemiologic Survey on Alcohol and Related Conditions (NESARC)	363	2001-2002
National Hospital Ambulatory Medical Care Survey (NHAMCS)	213 ER Dept. 305 Outpatient Dept.	2004
National Longitudinal Mortality Study (NLMS)	1,504	1979-1998
National Survey on Drug Use and Health (NSDUH)	218	2004
National Vital Statistics System: Natality (NVSS-N)	N/A	
Tobacco Use Supplement to the Current Population Survey (TUS-CPS)	222	November 2003

Table B-4. Sample Sizes for the NH/PI Race Category (continued)

Health Policy Issues (continued)	Number of NH/PI	Field Period
Washington State Population Survey (WSPS)	119 alone, 54 combined	2004
Youth Risk Behavior Surveillance Survey (YRBSS)	90	2005
Housing Issues	Number of NH/PI	Field Period
American Housing Survey (AHS)	139	2003
American Housing Survey: Metropolitan Surveys	122	2004
Justice System Issues	Number of NH/PI	Field Period
Annual Survey of Jails (ASJ)	N/A	
Census of Jails	N/A	
National Crime Victimization Survey (NCVS)	71	2003- 2004
Transportation	Number of NH/PI	Field Period
National Household Travel Survey (NHTS)	370 Head of HH 1,027 Person in HH 720 Primary drivers 3,794 Day travelers 127 Long-trip travelers	2001

Table B-5. Sample Sizes for the NH Alone Race Category

Demographic and Economic Indicators	Number of NH alone	Field Period
American Community Survey (ACS)	N/A	
Census 2000	140,652	2000
Elder Well-being	Number of NH only	Field Period
Health and Retirement Study (HRS)	N/A	1992
Education	Number of NH only	Field Period
Early Childhood Longitudinal Study, Birth Cohort (ECLS-B)	N/A	2001- 2002
Health Policy Issues	Number of NH only	Field Period
Hawaii Health Survey (HHS)	N/A	2004
National Vital Statistics System: Linked Birth-Infant Death (NVSS-I)	6,772 births	2002
National Vital Statistics System: Mortality (NVSS-M)	594	2003
National Vital Statistics System: Natality (NVSS-N)	N/A	
Pregnancy Risk Assessment Monitoring System (PRAMS)	N/A	
Military Service/Veterans Issues	Number of NH only	Field Period
National Survey of Veterans (NSV)	34	2001

Table B-6. Sample Sizes for the PI Alone Race Category

Demographic and Economic Indicators	Number of PI alone	Field Period
American Community Survey (ACS)	N/A	
Census 2000	58,240 Guamanian/Chamorro 91,029 Samoan 108,914 Other PI	2000
Elder Well-being	Number of PI only	Field Period
Health and Retirement Study (HRS)	N/A	1992
Education	Number of PI only	Field Period
Early Childhood Longitudinal Study, Birth Cohort (ECLS-B)	N/A	2001-2002
Health Policy Issues	Number of PI only	Field Period
National Vital Statistics System: Mortality (NVSS-M)	404 Samoan 159 Guamanian	2003
Military Service/Veterans Issues	Number of PI only	Field Period
National Survey of Veterans (NSV)	48	2001

Appendix C – List of Reports Reviewed

Literature and Reports Reviewed

- Barnes, P.M., Adams, P.F. and Powell-Griner, E., "Health Characteristics of the American Indian and Alaska Native Adult Population: United States, 1999-2003," *Advance Data from Vital and Health Statistics*, Number 356, April 27, 2005.
- National Research Council. *Eliminating Health Disparities: Measurement and Data Needs*. Panel on DHHS Collection of Race and Ethnicity Data, Michele Ver Ploeg and Edward Perrin, Editors. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- Finch, B.K., Morton, S., Elliot, M.N., Golinelli, D., Lurie, N., Do, D.P., Rastegar, A., Griffin, A.R., and Valentine, D., *Draft Report: Evaluation of Statistical Methods for Data Collection and Analysis on Racial and Ethnic Minorities and Other Hard-to-Reach Populations*. Contract No. 282-00-0005. Prepared for U.S. DHHS, Office of Minority Health, Rockville, MD: Undated.
- Indian Health Service, U.S. Department of Health and Human Services, Adjusting for Miscoding of Indian Race on State Death Certificates. Rockville, MD: 1996
- Moy, E., Smith, C.R., Johansson, P., and Andrews, R., "Gaps in Data for American Indians and Alaska Natives in the National Healthcare Disparities Report." *American Indian and Alaska Native Mental Health Research: The Journal of the National Center*. American Indian and Alaska Native Programs, University of Colorado at Denver and Health Sciences Center: 13 (1), 52-69. 2006.
- National Committee on Vital and Health Statistics, Subcommittee on Populations. Eliminating Health Disparities: Strengthening Data on Race, Ethnicity, and Primary Language in the United States. Hyattsville, MD: August 2005.
- National Committee on Vital and Health Statistics, Subcommittee on Populations, Minutes of Meeting on AI/AN Issues. Denver, CO, September 27, 2002.
- National Committee on Vital and Health Statistics, Subcommittee on Populations, Proceedings of Meeting on Asian, Native Hawaiians, and Other Pacific Islanders Data on Race/Ethnicity, May 22, 2003.
- National Committee on Vital and Health Statistics, Health Data Needs of the Pacific Insular Areas, Puerto Rico, and the U.S. Virgin Islands, Subcommittee on Populations, December, 1999
- Roberts, M. Racial and ethnic differences in health insurance coverage and usual source of health care, 2002. Rockville (MD): Agency for Healthcare Research and Quality; 2006. MEPS Chartbook No. 14. AHRQ Pub. No. 06-0004.
- Rhoades, D. "Commentary—Disparities in Data for American Indians and Alaska Natives," *American Indian and Alaska Native Mental Health Research: The Journal of the National Center*, American Indian and Alaska Native Programs, University of Colorado at Denver and Health Sciences Center: 13(1) 70-74. 2006.

- Strong, D.A., Del Grosso, P., Burwick, A., Jethwani, V., and Ponza, P., "Rural Research Needs and Data Sources for Selected Human Services Topics." Prepared for U.S. DHHS/ASPE: Mathematica Policy Research, Princeton, NJ, August 2005.
- U.S. Department of Health and Human Services, "Improving the Collection and Use of Racial and Ethnic Data in HHS," Joint Report of the DHHS Data Council's Working Group on Racial and Ethnic Data and the Data Work Group for the DHHS Initiative to Eliminate Racial and Ethnic Disparities in Health (undated).
- U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Office of Minority Health, "Highlights in Minority Health, Asian American and Pacific Islander Heritage Month." May 2005. (Accessed 9-6-05 at <http://www.cdc.gov/omb/Highlights/2005/HMary05.htm>).
- U.S. Department of Health and Human Services Data Council, Meeting Minutes, April-October 2005.
- U.S. Government Accountability Office, "Report on Implementation of the Indian Child Welfare Act," Washington, DC: April 8, 2005.
- Waksberg, J., Levine, D., and Marker, D. "Assessment of Major Federal Data Sets for Analyses of Hispanic and Asian or Pacific Islander Subgroups and Native American: Task 3 Report: Extending the Utility of Federal Databases." Prepared for DHHS/ASPE: Westat, Rockville, MD, May 2000.

Appendix D – Key Informant Interview Guide

INTERVIEW GUIDE

Brief Introduction/Overview of Data Study Objectives and Preliminary Findings

Purpose of this Interview

- **Perspectives on issues that contribute to current data gaps/limitations**
- **Input on current DHHS and other federal initiatives that are intended/would improve data on AI/AN/NA health and well-being**
- **Suggestions for other steps federal government or others could take to further improve data, including feasibility considerations**

Discussion Probes:

1. From your experience and knowledge of federal administrative data and surveys, what do you think are the major reasons why there are gaps in the data available on AI/AN/NA health and well-being?
2. Are you aware of changes that DHHS or other federal agencies are making (or are planned in the next few years) to administrative data sets or to ongoing and new federal surveys that would likely improve data on AI/AN/NA health and well-being?
3. What administrative or survey data sets within DHHS would you say should be the highest priority for improving AI/AN/NA data?
4. Are there other suggestions you may have, or that others have made, that would improve data on AI/AN/NA health and well-being?
5. Are there other people within DHHS or other federal agencies that you would recommend we talk with about these issues?