

Free Executive Summary

Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement

Cheryl Ulmer, Bernadette McFadden, and David R. Nerenz, Editors; Subcommittee on Standardized Collection of Race/Ethnicity Data for Healthcare Quality Improvement; Institute of Medicine

ISBN: 978-0-309-14012-6, 204 pages, 8.5 X 11, paperback (2009)



This free executive summary is provided by the National Academies as part of our mission to educate the world on issues of science, engineering, and health. If you are interested in reading the full book, please visit us online at <http://www.nap.edu/catalog/12696.html>. You may browse and search the full, authoritative version for free; you may also purchase a print or electronic version of the book. If you have questions or just want more information about the books published by the National Academies Press, please contact our customer service department toll-free at 888-624-8373.

This executive summary plus thousands more available at www.nap.edu.

Copyright © National Academy of Sciences. All rights reserved. Unless otherwise indicated, all materials in this PDF file are copyrighted by the National Academy of Sciences. Distribution or copying is strictly prohibited without permission of the National Academies Press <http://www.nap.edu/permissions/>. Permission is granted for this material to be posted on a secure password-protected Web site. The content may not be posted on a public Web site.

Summary

The goal of eliminating disparities in health care in the United States remains elusive. The findings of the *National Healthcare Disparities Reports* reveal that even as quality improves on specific measures, disparities often persist (AHRQ, 2008a, 2008b). Addressing these disparities must begin with the fundamental step of bringing the nature of the disparities and the groups at risk for those disparities to light by collecting health care quality information stratified by race, ethnicity and language data. Then attention can be focused on where interventions might be best applied, and on planning and evaluating those efforts to inform the development of policy and the application of resources. A lack of standardization of categories for race, ethnicity, and language data has been suggested as one obstacle to achieving more widespread collection and utilization of these data. Many types of entities participate in initiatives to improve the quality of health care; health plans, hospitals, other providers, and health systems can and should obtain race, ethnicity, and language data so these data can be used to identify gaps and improve care for all individuals.

The purpose of this report is to identify standardized categories for the variables of race, ethnicity, and language that can be used to facilitate the sharing, compilation, and comparison of quality data stratified by the standard categories. The Institute of Medicine, under a contract with the Agency for Healthcare Research and Quality (AHRQ), Department of Health and Human Services (HHS) formed the Subcommittee on Standardized Collection of Race/Ethnicity Data for Healthcare Quality Improvement to identify current models for collecting and coding race, ethnicity, and language data; to ascertain the challenges involved in obtaining these data in health care settings; and to make recommendations for improvement. The language in the statement of task (Box S-1)—“in healthcare quality improvement” and “assess and report on quality of care”—led the subcommittee to focus its discussion and recommendations on data collection in the domain of health care services.

EXISTING GUIDANCE ON RACE, ETHNICITY, AND LANGUAGE CATEGORIES

The concepts of race and ethnicity are defined socially and culturally and, in the case of federal data collection, by legislative and political necessity (Hayes-Bautista and Chapa, 1987). With the aim of identifying important cultural and social groups for statistical reporting and civil rights monitoring, the Office of Management and Budget (OMB) has developed a minimum set of standardized categories for reporting on race and Hispanic ethnicity by federal agencies and recipients of federal funds (OMB, 1977, 1997b). The five race categories are now Black or African American, White, Asian, American Indian or Alaska Native, and Native Hawaiian or Other Pacific Islander.

BOX S-1
**Statement of Task: Subcommittee on Standardized Collection of
Race/Ethnicity Data for Healthcare Quality Improvement**

A subcommittee of experts will report to the IOM Committee on Future Directions for the National Healthcare Quality and Disparities Reports regarding the lack of standardization of collection of race and ethnicity data at the federal, state, local, and private sector levels due to the fact that the federal government has yet to issue comprehensive, definitive guidelines for the collection and disclosure of race and ethnicity data in healthcare quality improvement. The subcommittee will focus on defining a standard set of race/ethnicity and language categories and methods for obtaining this information to serve as a standard for those entities wishing to assess and report on quality of care across these categories. The subcommittee will carry out an appropriate level of detailed, in-depth analysis and description which can be included in the overall report by the committee and as a separate stand alone report.

OMB describes these categories as the minimum set and encourages the collection of more detailed data provided those data can be aggregated back to the minimum categories (OMB, 1997a). Progress has been made in incorporating these categories into the collection and presentation of data in health care settings. However, some health care–related data collection efforts still do not employ these basic standard categories.

While OMB has not established a list of language categories, the collection of language data has been pivotal in determining whether there has been discrimination by “national origin” under Title VI of the Civil Rights Act of 1964,^{1,2} and federal policies state that “reasonable steps” need to be taken so that persons of limited English proficiency can have “meaningful access” to programs or activities without charge for language services.³ Additionally in 2000, HHS released its National Standards on Culturally and Linguistically Appropriate Services (CLAS), which encourage all health care organizations and individual providers “to make their practices more culturally and linguistically accessible,” including the use of race, ethnicity, and language data in program assessments and incorporation of these data into health records and organizational management systems (HHS, 2007).

CATEGORIZING RACE AND ETHNICITY DATA

The OMB race and Hispanic ethnicity categories represent broad population groups used for an array of statistical reporting and analytic purposes, including health care quality assessment and identification of disparities (AHRQ, 2008a; Cohen, 2008; Flores and Tomany-Korman, 2008; IOM, 2008; Kaiser Family Foundation, 2009). Chapter 2 illustrates that these categories alone, however, are insufficient to illuminate many disparities and to target quality improvement efforts where they may be most needed. Since disparities can exist within those broad OMB categories, there is value in collecting and utilizing data incorporating more fine-grained categories than those of OMB (Blendon et al., 2007; Jerant et al., 2008; Read et al., 2005; Shah and Carrasquillo, 2006). The subcommittee recommends a separate question to collect data on granular ethnicity—defined as “a person’s ethnic origin or descent, ‘roots,’ or heritage, or the place of birth of the person or the person’s parents or ancestors...” (U.S. Census Bureau, 2008)—in addition to soliciting data in the OMB race and Hispanic ethnicity categories (Figure S-1). Research also shows that not all individuals identify with the current OMB race categories so the

¹ *The Civil Rights Act of 1964*, Public Law 88-352, 78 Stat. 241, 88th Cong., 2nd sess. (July 2, 1964).

² *Lau v Nichols*, 414 U.S. 563 (1974).

³ *Improving Access to Services for Persons with Limited English Proficiency*, Executive Order 13166, August 11, 2000.

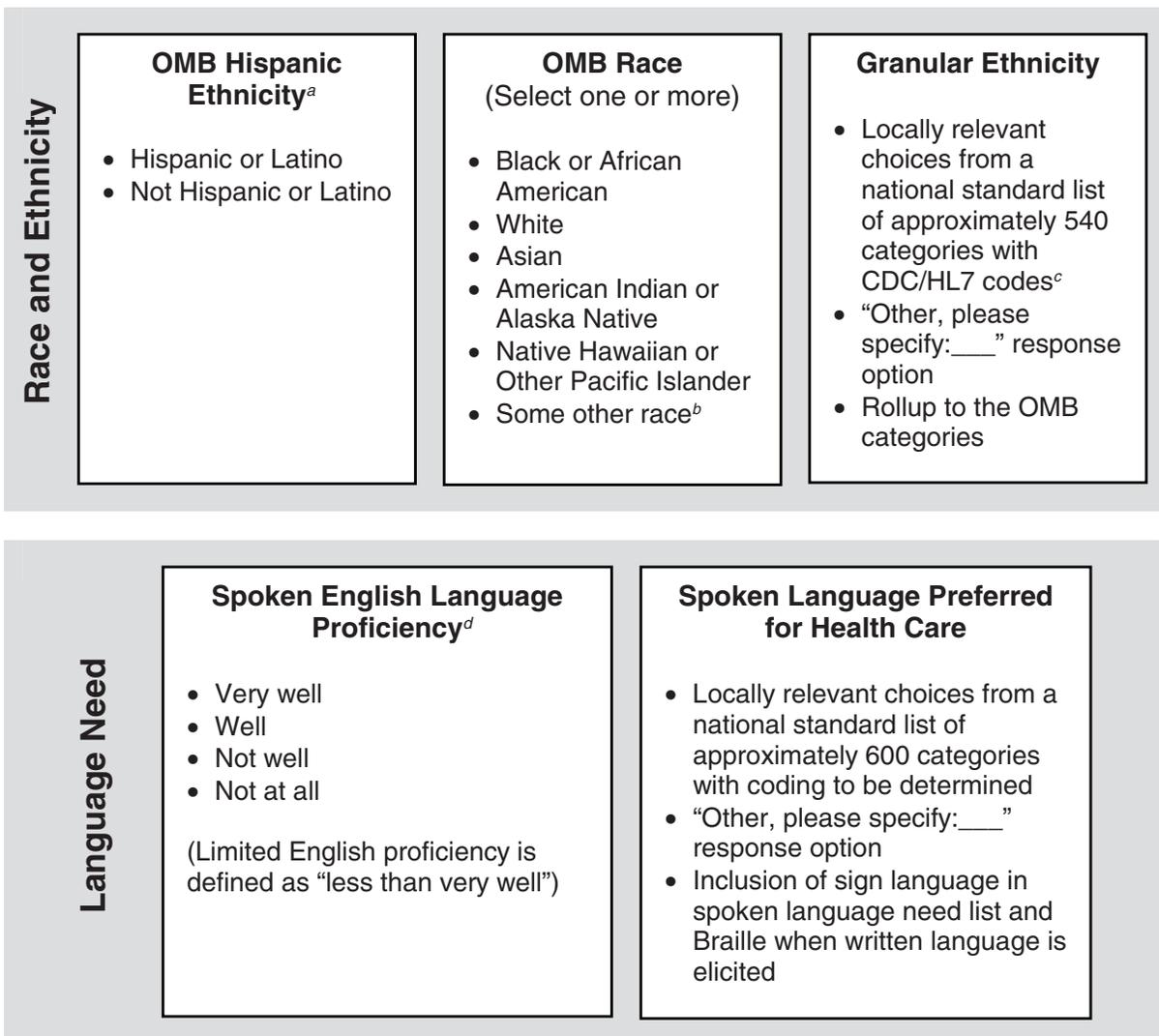


FIGURE S-1 Recommended variables for standardized collection of race, ethnicity, and language need.

^a The preferred order of questioning is Hispanic ethnicity first, followed by race, and then granular ethnicity.

^b The U.S. Census Bureau received OMB permission to add “Some other race” to the standard OMB categories in Census 2000 and subsequent Census collections.

^c Additional codes will be needed for categories added to the CDC/HL7 list.

^d Need is determined on the basis of two questions, asking about proficiency first. Limited English proficiency is defined for health care purposes as speaking English less than very well.

Note: Additional categories for HIT tracking might include whether respondents have not yet responded (unavailable), refuse to answer (declined), or do not know (unknown), as well as whether responses are self-reported or observer-reported.

Sources: CDC, 2000; Office of Management and Budget, 1997b; U.S. Census Bureau, 2002; Shin and Bruno, 2003.

subcommittee recommends expanding the race categories to six choices by including a “Some other race” option to provide a response category for those Hispanics and others who do not relate to the current choices.

In Chapter 3, the subcommittee considers whether a national “OMB Plus” set of 10 to 15 granular ethnicity categories, similar to the Census Bureau approach, should be identified that would be optimal for collection by all health care entities. However, such a set would not be specific to and appropriate for the diverse communities in which health care entities operate. Instead, the subcommittee concludes that individual entities should select the granular ethnicity categories representative of their service population selected from a national list of standardized categories. Whenever a limited list of categories is offered to respondents, the list should include an open-ended response option of “Other, please specify:___” so that each individual who desires to do so can self-identify.

Recommendation 3-1: An entity collecting data from individuals for purposes related to health and health care should:

- **Collect data on granular ethnicity using categories that are applicable to the populations it serves or studies. Categories should be selected from a national standard list (see Recommendation 6-1a) on the basis of health and health care quality issues, evidence or likelihood of disparities, or size of subgroups within the population. The selection of categories should also be informed by analysis of relevant data (e.g., Census data) on the service or study population. In addition, an open-ended option of “Other, please specify:___” should be provided for persons whose granular ethnicity is not listed as a response option.**
- **Elicit categorical responses consistent with the current OMB standard race and Hispanic ethnicity categories, with the addition of a response option of “Some other race” for persons who do not identify with the OMB race categories.**

While several organizations provide lists of granular ethnicities (e.g., Centers for Disease Control and Prevention [CDC]/Health Level 7 [HL7] and the Commonwealth of Massachusetts/Brookings Institution), none of these lists is sufficient for a standard national set from which locally relevant choices could be made (CDC, 2000; Taylor-Clark et al., 2009). A merged list provides a template from which such a national standard set can be developed (see Appendix E). When a person does not check off an OMB race or Hispanic ethnicity and provides only a granular ethnicity response, a process for rolling granular ethnicity categories up to the OMB categories will, in some cases, be necessary for analysis and reporting purposes. However, some ethnicities do not correspond to a single OMB race category, necessitating a “no determinate OMB classification” for analytic purposes (see Appendix F).

Recommendation 3-2: Any entity collecting data from individuals for purposes related to health and health care should collect granular ethnicity data in addition to data in the OMB race and Hispanic ethnicity categories and should select the granular ethnicity categories to be used from a national standard set. When respondents do not self-identify as one of the OMB race categories or do not respond to the Hispanic ethnicity question, a national scheme should be used to roll up the granular ethnicity categories to the applicable broad OMB race and Hispanic ethnicity categories to the extent feasible.

Eliciting accurate and reliable race, Hispanic ethnicity, and granular ethnicity data depends on the ways in which the questions are asked, the instructions provided to respondents (e.g., “Select one or more”), and the format of the questions (i.e., OMB one-question versus two-question format). This latter issue is especially relevant to how Hispanic populations self-identify. Pilot projects and further study are necessary to confirm the best ways to collect accurate data that are useful for health care quality improvement.

Recommendation 3-3: To determine the utility for health and health care purposes, HHS should pursue studies on different ways of framing the questions and related response categories for collecting race and ethnicity data at the level of the OMB categories, focusing on completeness and accuracy of response among all groups.

- **Issues addressed should include use of the one- or two-question format for race and Hispanic ethnicity, whether all individuals understand and identify with the OMB race and Hispanic ethnicity categories, and the increasing size of populations identifying with “Some other race.”**
- **The results of such studies, together with parallel studies by the Census Bureau and other agencies, may reveal the need for an OMB review across all agencies to determine the best format for improving response among all groups.**

IMPROVING THE COLLECTION OF DATA ON LANGUAGE

Compelling evidence exists that having limited English proficiency (LEP) affects the delivery and quality of health care and can result in significant disparities in access to care (Hu and Covell, 1986; Weinick and Krauss, 2000), a decreased likelihood of having a usual source of care (Kirkman-Liff and Mondragon, 1991; Weinick and Krauss, 2000), an increased probability of receiving unnecessary diagnostic tests (Hampers et al., 1999), more serious adverse outcomes from medical errors (Divi et al., 2007), and more drug-related complications (Gandhi et al., 2000). To achieve safe, effective, patient-centered communication, attention must be paid to the language needs of patients, as addressed in Chapter 4.

Language Questions

Assessing each individual’s language need is an essential first step toward ensuring effective health care communication. The subcommittee concludes that spoken language need can best be assessed by asking two questions: one aimed at determining whether an individual speaks English less than very well and a second aimed at identifying the individual’s preferred spoken language during a health care encounter (Figure S-1). Having this information for each individual allows its use to ensure the quality of services in subsequent encounters, in analysis of health care disparities, and in system-level planning (e.g., determining the need for interpreters and matching patients to language-concordant providers).

The subcommittee establishes a hierarchy among the possible language questions, with questions about English proficiency and preferred spoken language identified as a higher priority than questions on language spoken at home or on preferred language for written materials. On average, 55 percent of those who speak another language at home speak English very well (Shin and Bruno, 2003), but asking about language spoken at home helps provide a window into the health beliefs and practices of the home environment. The correlation between those who need spoken and written language assistance appears to be high in many settings.

Recommendation 4-1: To assess patient/consumer language and communication needs, all entities collecting data from individuals for purposes related to health and health care should:

- **At a minimum, collect data on an individual’s assessment of his/her level of English proficiency and on the preferred spoken language needed for effective communication with health care providers. For health care purposes, a rating of spoken English-language proficiency of less than very well is considered limited English proficiency.**
- **Where possible and applicable, additionally collect data on the language spoken by the individual at home and the language in which he/she prefers to receive written materials.**

When the individual is a child, the language need of the parent/guardian must be determined. Similarly, if an adult has a guardian/conservator, that language information must be assessed.

Languages in Use

More than 600 languages are in use in the United States. In Chapter 4 the subcommittee evaluates options for determining what language categories entities should use for data collection (e.g., a uniform set for all entities,

percentage or numerical thresholds based on the presence of languages in a service area, or local choice). Local choice informed by data on the languages spoken most frequently in the service area by persons with LEP is the preferred option. A single list does not suit all areas given that the top non-English languages vary greatly from area to area (for instance, Spanish is in the top 10 languages in 3,122 of 3,141 counties in the United States, while Turkish is in the top 10 in 12 counties, Laotian in 125, Navaho in 74, SerboCroatian in 58, and Portuguese in 229) (U.S. English Foundation, 2009). The aim is to have data on each individual's specific language need, but when an entity designs its collection instruments, whether paper or electronic, it may, because of space considerations, have to use a limited number of response categories. Therefore, such a response list should always include an "Other, please specify: __" option. Some electronic data collection systems are more sophisticated, and by using keystroke recognition can accommodate hundreds of languages.

Recommendation 4-2: The choice of response categories for spoken and written language questions should be informed by analysis of relevant data on the service area (e.g., Census data) or service population, and any response list should include an option of "Other, please specify: __" for persons whose language is not listed.

The subcommittee has developed a template of languages used in the United States based on Census data and the experiences of certain health care providers. This template can serve as a basis for the national standardized set called for in recommendations in Chapter 6 (see Appendix I for template). A uniform set of codes can facilitate sharing of data. Two possible language coding systems already exist (the Census and International Organization for Standardization [ISO] code sets) (Census 2002a; SIL International, 2009).

Recommendation 4-3: When any health care entity collects language data, the languages used as response options or categories for analysis should be selected from a national standard set of languages in use in the United States. The national standard set should include sign language(s) for spoken language and Braille for written language.

IMPROVING DATA COLLECTION ACROSS THE HEALTH CARE SYSTEM

As discussed in Chapter 5, while each of the entities involved in the nation's health care system has some capability for the collection of race, ethnicity, and language data, some are better positioned than others to collect these data through self-report, the generally agreed-upon best way to define a person's racial and ethnic identity. In the future, information infrastructure may enable integrated data exchange so that all entities will not need to collect all data. For now, however, all health and health care entities have roles to play in collecting these data directly from individuals. Hospitals, community health centers, physician practices, health plans, and local, state, and federal agencies can all identify next steps toward improving or implementing direct data collection by understanding the unique contexts in which they operate. Across all these entities, these data must be collected and stored responsibly. Training of staff, upgrades to health information technology (HIT) systems, and communication with patients and enrollees are potential avenues for improved data collection and building of trust.

In the subcommittee's proposed framework, optional categories are offered (e.g., declined, unavailable, unknown, self-reported, observer reported); these are not for patient response, but for tracking the portion of the patient population for which an entity has been able to collect data or the nature of the data collection. Until directly collected data are sufficient for analytic and quality improvement purposes, indirect estimation of race and ethnicity through techniques such as geocoding and surname analysis is useful for bridging data gaps.

Recommendation 5-1: Where directly collected race and ethnicity data are not available, entities should use indirect estimation to aid in the analysis of racial and ethnic disparities and in the development of targeted quality improvement strategies, recognizing the probabilistic and fallible nature of such indirectly estimated identifications.

- **Race and ethnicity identifications based on indirect estimation should be distinguished from self-reports in data systems, and if feasible, should be accompanied by probabilities.**
- **Interventions and communications in which race and ethnicity identifications are based on indirect estimation may be better suited to population-level interventions and communications and less well suited to use in individual-level interactions.**
- **An indirectly estimated probability of an individual's race and ethnicity should never be placed in a medical record or used in clinical decision making.**
- **Analyses using indirectly estimated race and ethnicity should employ statistically valid methods that deal with probabilistic identifications.**

IMPLEMENTING COLLECTION OF STANDARDIZED DATA

Now is an opportune time for action on standardization of the categories used to collect race, ethnicity, and language data. Moreover, efforts to share and evaluate quality data across states, regions, or payers would be facilitated by standardized categories.

HHS is a prime locus of the subcommittee's recommendations in Chapter 6 for implementation of improved collection of standardized data because of its focus on resolving health and health care disparities and its history of promoting the collection of race, ethnicity, and language data to ensure compliance with applicable statutes and regulations. National development of standardized categories and coding by HHS, along with a responsive updating process, would relieve each state and entity of having to develop its own set of categories and coding scheme, which could be incompatible with others. The subcommittee templates of categories along with an updated CDC/HL7 code set can form the basis for standardized race, Hispanic ethnicity, and granular ethnicity data while a determination will have to be made on coding for languages.

Recommendation 6-1a: HHS should develop and make available national standard lists of granular ethnicity categories and spoken and written languages, with accompanying unique codes and rules for rollup procedures.

- **HHS should adopt a process for routine updating of those lists and procedures as necessary. Sign languages should be included in national lists of spoken languages and Braille in lists of written languages.**
- **HHS should ensure that any national hierarchy used to roll up granular ethnicity categories to the broad OMB race and Hispanic ethnicity categories takes into account responses that do not correspond to one of the OMB categories.**

Standardization would support achievement of the goal set forth in the American Recovery and Reinvestment Act of 2009⁴ (ARRA) of having a national electronic health record (EHR)⁵ for each individual by 2014, incorporating collection of data on the person's race, ethnicity, and primary language. Having the standards adopted by the other components of the health care industry, including the makers of HIT systems, would help ensure that a sufficient set of data fields are available to accommodate each element recommended for collection by the subcommittee.

Recommendation 6-1b: HHS and the Office of the National Coordinator for Health Information Technology (ONC) should adopt standards for including in electronic health records the variables of race, Hispanic ethnicity, granular ethnicity, and language need identified in this report.

⁴ *American Recovery and Reinvestment Act of 2009*, Public Law 111-5 § 3002(b)(2)(B)(vii), 111th Cong., 1st sess. (February 17, 2009)

⁵ In this document, EHR means a patient record owned and maintained by a provider entity; a personal health record is a medical or health record owned and maintained by a patient him- or herself.

Recommendation 6-1c: HHS and ONC should develop standards for electronic data transmission among health care providers and plans that support data exchange and possible aggregation of race, Hispanic ethnicity, granular ethnicity, and language need data across entities to minimize redundancy in data collection.

Performance incentive programs tend not to be designed with reduction of disparities in mind, yet can have positive or negative effects on disparities in health care and on underresourced primary care safety net providers (Chien et al., 2007; Rust and Cooper, 2007; Williams, 2009). The subcommittee does not take a stand on whether incentive payments in HIT programs should exist, but when they do exist, the collection of race, ethnicity, and language data would be one activity for which positive incentives should be offered.

Recommendation 6-1d: The Centers for Medicare and Medicaid Services (CMS), as well as others sponsoring payment incentive programs, should ensure that the awarding of such incentives takes into account collection of the recommended data on race, Hispanic ethnicity, granular ethnicity, and language need so these data can be used to identify and address disparities in care.

Numerous past and present legislative and policy efforts stress the importance of collecting race, ethnicity, and language data in federal programs. HHS administers programs supporting the health care delivery system to provide care to persons at risk of receiving suboptimal care, and these programs present opportunities to influence the quality of care delivered to millions of Americans. Because the subcommittee's charge relates to health care, the following recommendation focuses on the HHS programs that deliver health care services, pay for those services through insurance mechanisms, or administer surveys that increase knowledge on health care needs and outcomes. The Secretary, however, may find it useful to extend the standardized approach of this report to other HHS health-related programs or other data gathering activities.

Recommendation 6-1e: HHS should issue guidance that recipients of HHS funding (e.g., Medicare, the Children's Health Insurance Program [CHIP], Medicaid, community health centers) include data on race, Hispanic ethnicity, granular ethnicity, and language need in individual health records so these data can be used to stratify quality performance metrics, organize quality improvement and disparity reduction initiatives, and report on progress.

Having quality-of-care information from large federal delivery systems such as the Department of Veterans Affairs, the Department of Defense, and other federally funded programs, such as community health centers, stratified by the same variables and categories recommended in this report would provide rich sources for comparative analysis.

Recommendation 6-2: HHS, the Department of Veterans Affairs, and the Department of Defense should coordinate their efforts to ensure that all federally funded health care delivery systems collect the variables of race, Hispanic ethnicity, granular ethnicity, and language need as outlined in this report, and include these data in the health records of individuals for use in stratifying quality performance metrics, organizing quality improvement and disparity reduction initiatives, and reporting on progress.

Accreditation organizations and other professional and standards-setting bodies can play a key role in fostering the collection of race, ethnicity, and language data. Hospitals, health plans, and physicians have reported that a lack of standardization has been a barrier to using these data in quality improvement efforts (Bilheimer and Sisk, 2008; Lurie et al., 2008; NCQA, 2009; Siegel et al., 2008). The Joint Commission, the National Committee for Quality Assurance (NCQA), and URAC⁶ have developed CLAS-like standards for their organizational reviews.

⁶ Formerly known as the Utilization Review Accreditation Commission.

The National Quality Forum (NQF) encourages the collection of race, ethnicity, and language data in accordance with the Health Research & Educational Trust (HRET) Toolkit (NQF, 2008); the subcommittee's recommendations include modifications to that toolkit. The American Medical Association, the National Medical Association, and the National Hispanic Medical Association's Commission to End Health Care Disparities have reaffirmed their collective commitment to bringing an end to health care disparities by increasing awareness in the physician community and promoting better data collection (AMA, 2005, 2009).

Recommendation 6-3: Accreditation and standards-setting organizations should incorporate the variables of race, Hispanic ethnicity, granular ethnicity, and language need outlined in this report and associated categories (as updated by HHS) as part of their accreditation standards and performance measure endorsements.

- **The Joint Commission, NCQA, and URAC should ensure collection in individual health records of the variables of race, Hispanic ethnicity, granular ethnicity, and language need as outlined in this report so these data can be used to stratify quality performance metrics, organize quality improvement and disparity reduction initiatives, and report on progress.**
- **NQF should review and amend its recommendations on the collection and use of data on race, Hispanic ethnicity, granular ethnicity, and language need to accord with the categories and procedures outlined in this report.**
- **Medical societies and medical boards should review and endorse the variables, categories, and procedures outlined in this report and educate their members on their use for quality improvement.**

States have an opportunity to shape the level of detail for race, ethnicity, and language data collected in their programs whether for use in reporting on quality measures by insurance programs, in disease registries, in hospital discharges, in health care surveys, in patient safety reporting, or in other activities. Through Medicaid and CHIP programs, states have leverage with managed care organizations and providers to require collection of the recommended data and their use in quality improvement. Medicaid provides coverage for a large portion of minority groups, and states have an interest in ensuring that the population covered is receiving appropriate quality care (Angeles and Somers, 2007).

Recommendation 6-4: Through their certification, regulation, and monitoring of health care providers and organizations within their jurisdiction, states should require the collection of data on the race, Hispanic ethnicity, granular ethnicity, and language need variables as outlined in this report so these data can be used to stratify quality performance metrics, organize quality improvement and disparity reduction initiatives, and report on progress.

CONCLUSION

Efforts are under way to establish national standards for health care technology, performance measurement, and data aggregation and exchange that complement local data collection and experiences with performance improvement and reporting (Roski, 2009). To date, it has been difficult to either combine or compare performance data stratified by race, ethnicity, or language need across payment and delivery systems, which has limited the utility of such data for assessing the performance of the health system as a whole or in specific geographic regions with respect to disparities. Yet, these analyses have implications for the design of appropriate interventions by federal, state, and local policy makers and health care plans and providers.

Standardization of the categories used to collect these data would promote greater comparability of patient-focused data collected directly by care providers or health plans, or, for instance, transferred from providers to multiple plans. Standardization would also eliminate the need for all health care entities to develop their own categorization schemes. Still, additional resources and leadership at the local, state, and national levels will be

required to implement these recommendations. Although broad application of EHRs will take a number of years, the data collection issues for current systems do not differ significantly from those involved in future EHR applications, so providers could institute today the processes for the capture and sharing of race, ethnicity, and language data proposed in this report.

There is strong evidence that the quality of health care varies by race, ethnicity, and language. Quality metrics stratified by race, Hispanic ethnicity, granular ethnicity and language need can inform point-of-care services, application of resources, and decisions in patient–provider interactions in ways that can assist in improving absolute quality and reducing disparities.

REFERENCES

- AHRQ (Agency for Healthcare Research and Quality). 2008a. *National Healthcare Disparities Report*. Rockville, MD: AHRQ.
- . 2008b. *The National Healthcare Quality Report*. Rockville, MD: AHRQ.
- AMA (American Medical Association). 2005. *Opinion 9.121 - racial and ethnic health care disparities*. <http://www.ama-assn.org/ama/pub/physician-resources/medical-ethics/code-medical-ethics/opinion9121.shtml> (accessed May 22 2009).
- . 2009. *Commission to End Health Care Disparities, five year summary*. Chicago, IL: American Medical Association.
- Angeles, J., and S. A. Somers. 2007. *From policy to action: Addressing racial and ethnic disparities at the ground-level*. Hamilton, NJ: Center for Health Care Strategies, Inc.
- Bilheimer, L. T., and J. E. Sisk. 2008. Collecting adequate data on racial and ethnic disparities in health: The challenges continue. *Health Affairs* 27:383-391.
- Blendon, R. J., T. Buhr, E. F. Cassidy, D. J. Perez, K. A. Hunt, C. Fleischfresser, J. M. Benson, and M. J. Herrmann. 2007. Disparities in health: Perspectives of a multi-ethnic, multi-racial America. *Health Affairs* 26(5):1437-1447.
- CDC (Centers for Disease Control and Prevention). 2000. *Race and ethnicity Code set version 1.0*. Atlanta, GA: Centers for Disease Control and Prevention.
- Chien, A. T., M. H. Chin, A. M. Davis, and L. P. Casalino. 2007. Pay for performance, public reporting, and racial disparities in health care: How are programs being designed? *Medical Care Research and Review* 64(5 Suppl):283S-304S.
- Cohen, L. L. 2008. Racial/ethnic disparities in hospice care: A systematic review. *Journal of Palliative Medicine* 11(5):763-768.
- Divi, C., R. G. Koss, S. P. Schmaltz, and J. M. Loeb. 2007. Language proficiency and adverse events in US hospitals: A pilot study. *International Journal for Quality in Health Care* 19(2):60-67.
- Flores, G., and S. C. Tomany-Korman. 2008. Racial and ethnic disparities in medical and dental health, access to care, and use of services in US children. *Pediatrics* 121(2):e286-e298.
- Gandhi, T. K., H. R. Burstin, E. F. Cook, A. L. Puopolo, J. S. Haas, T. A. Brennan, and D. W. Bates. 2000. Drug complications in outpatients. *Journal of General Internal Medicine* 15(3):149-154.
- Hampers, L. C., S. Cha, D. J. Gutglass, H. J. Binns, and S. E. Krug. 1999. Language barriers and resource utilization in a pediatric emergency department. *Pediatrics* 103(6 Pt 1):1253-1256.
- Hayes-Bautista, D. E., and J. Chapa. 1987. Latino terminology: Conceptual bases for standardized terminology. *American Journal of Public Health* 77:61-68.
- HHS (U.S. Department of Health and Human Services), Office of Minority Health (OMH). 2007. *National standards on culturally and linguistically appropriate services (CLAS)*. <http://www.omhrc.gov/templates/browse.aspx?lvl=2&lvlID=15> (accessed May 13, 2009).
- Hu, D. J., and R. M. Covell. 1986. Health care usage by Hispanic outpatients as function of primary language. *Western Journal of Medicine* 144(4):490-493.
- IOM (Institute of Medicine). 2008. *Challenges and successes in reducing health disparities: Workshop summary*. Washington, DC: The National Academies Press.
- Jerant, A., R. Arellanes, and P. Franks. 2008. Health status among US Hispanics: Ethnic variation, nativity, and language moderation. *Medical Care* 46(7):709-717.
- Kaiser Family Foundation. 2009. *Putting women's health care disparities on the map: Examining racial and ethnic disparities at the state level*. Menlo Park, CA: The Henry J. Kaiser Family Foundation.
- Kirkman-Liff, B., and D. Mondragon. 1991. Language of interview: Relevance for research of Southwest Hispanics. *American Journal of Public Health* 81(11):1399-1404.
- Lurie, N., A. Fremont, S. A. Somen, K. Coltin, A. Gelzer, R. Johnson, W. Rawlins, G. Ting, W. Wong, and D. Zimmerman. 2008. The National Health Plan Collaborative to reduce disparities and improve quality. *Joint Commission Journal on Quality and Patient Safety* 34(5):256-265.
- NCQA (National Committee for Quality Assurance). 2009. *Supporting small practices: Lessons for health reform*. Washington, DC: NCQA.
- NQF (National Quality Forum). 2008. *National voluntary consensus standards for ambulatory care—measuring healthcare disparities*. Washington, DC: National Quality Forum.
- OMB (Office of Management and Budget). 1977. *Statistical policy directive No. 15, race and ethnic standards for federal statistics and administrative reporting*. <http://wonder.cdc.gov/wonder/help/populations/bridged-race/Directive15.html> (accessed August 3, 2009).

- . 1997a. Recommendations from the Interagency Committee for the Review of the Racial and Ethnic Standards to the Office of Management and Budget concerning changes to the standards for the classification of federal data on race and ethnicity. *Federal Register* (3110-01):36873-36946.
- . 1997b. Revisions to the standards for the classification of federal data on race and ethnicity. *Federal Register* 62:58781-58790.
- Read, J. G., B. Amick, and K. M. Donato. 2005. Arab immigrants: A new case for ethnicity and health? *Social Science and Medicine* 61(1):77-82.
- Roski, J. 2009. *Road map for better performance information through distributed data network*. Quality Alliance Steering Committee. Presentation to the IOM Committee on Future Directions for the National Healthcare Quality and Disparities Reports, February 10, 2009. Washington, DC. PowerPoint Presentation.
- Rust, G., and L. A. Cooper. 2007. How can practice-based research contribute to the elimination of health disparities? *Journal of the American Board of Family Medicine* 20(2):105-114.
- Shah, N. S., and O. Carrasquillo. 2006. Twelve-year trends in health insurance coverage among Latinos, by subgroup and immigration status. *Health Affairs* 25(6):1612-1619.
- Shin, H. B., and R. Bruno. 2003. *Language use and English-speaking ability: 2000*. Washington, DC: U.S. Census Bureau.
- SIL International. 2009. *Relationship between ISO 639-3 and the other parts of ISO 639*. <http://www.sil.org/iso639-3/relationship.asp> (accessed July 20, 2009).
- Siegel, B., J. Bretsch, K. Jones, V. Sears, L. Vaquerano, and M. J. Wilson. 2008. *Expecting success: Excellence in cardiac care results from Robert Wood Johnson Foundation quality improvement collaborative*. Princeton, NJ: Robert Wood Johnson Foundation.
- Taylor-Clark, K., A. B. Anise, Y. Joo, and M. Chin. 2009. *Massachusetts Superset*. Washington, DC: The Brookings Institution.
- U.S. Census Bureau. 2002a. *Census 2000 summary file 3: Technical documentation*. <http://www.census.gov/prod/cen2000/doc/sf3.pdf> (accessed August 3, 2009).
- U.S. Census Bureau. 2002b. *Modified race data summary file: 2000 Census of population and housing, technical documentation* <http://www.census.gov/popest/archives/files/MRSF-01-US1.html#fig1> (accessed February 25, 2009).
- . 2008. *Ancestry*. <http://www.census.gov/population/www/ancestry/ancoverview.html> (accessed May 24, 2009).
- U.S. English Foundation. 2009. *Languages in America*. <http://www.usefoundation.org/view/29> (accessed May 26, 2009).
- Weinick, R. M., and N. A. Krauss. 2000. Racial/ethnic differences in children's access to care. *American Journal of Public Health* 90(11):1771-1774.
- Williams, T. 2009. *Healthcare quality and disparities: Implications for pay for performance*. Integrated Health Association. Presentation to the IOM Committee on Future Directions for the National Healthcare Quality and Disparities Reports, March 12, 2009. Newport Beach, CA. PowerPoint Presentation.

RACE, ETHNICITY, AND LANGUAGE DATA

STANDARDIZATION FOR HEALTH CARE QUALITY IMPROVEMENT

Subcommittee on Standardized Collection of Race/Ethnicity Data for Healthcare Quality Improvement

Board on Health Care Services

Cheryl Ulmer, Bernadette McFadden, and David R. Nerenz, *Editors*

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

PREPUBLICATION COPY—Uncorrected Proofs

Copyright National Academy of Sciences. All rights reserved.
This executive summary plus thousands more available at <http://www.nap.edu>

THE NATIONAL ACADEMIES PRESS

500 Fifth Street, N.W.

Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This study was supported by Task No. #HHSP233200800005T between the National Academy of Sciences and the Agency for Healthcare Research and Quality. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the organizations or agencies that provided support for this project.

International Standard Book Number 0-309-XXXXXX-X (Book)

International Standard Book Number 0-309-XXXXXX-X (PDF)

Additional copies of this report are available from the National Academies Press, 500 Fifth Street, N.W., Lockbox 285, Washington, DC 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, <http://www.nap.edu>.

For more information about the Institute of Medicine, visit the IOM home page at: www.iom.edu.

Copyright 2009 by the National Academy of Sciences. All rights reserved.

Printed in the United States of America

The serpent has been a symbol of long life, healing, and knowledge among almost all cultures and religions since the beginning of recorded history. The serpent adopted as a logotype by the Institute of Medicine is a relief carving from ancient Greece, now held by the Staatliche Museen in Berlin.

Suggested citation: IOM (Institute of Medicine). 2009. *Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement*. Washington, DC: The National Academies Press.

*“Knowing is not enough; we must apply.
Willing is not enough; we must do.”*
—Goethe



INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

Advising the Nation. Improving Health.

PREPUBLICATION COPY—Uncorrected Proofs

Copyright National Academy of Sciences. All rights reserved.
This executive summary plus thousands more available at <http://www.nap.edu>

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Charles M. Vest is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

**SUBCOMMITTEE ON STANDARDIZED COLLECTION OF RACE/
ETHNICITY DATA FOR HEALTHCARE QUALITY IMPROVEMENT**

DAVID R. NERENZ (*Chair*), Director, Center for Health Services Research, Henry Ford Health System, Detroit, MI

MARGARITA ALEGRÍA, Professor, Department of Psychiatry, Harvard Medical School, and Director, Center for Multicultural Mental Health Research, Cambridge Health Alliance, Boston, MA

JOHN Z. AYANIAN, Professor of Medicine and Health Care Policy, Harvard Medical School and Brigham and Women's Hospital, Boston, MA

IGNATIUS BAU, Program Director, The California Endowment, Oakland, CA

RODERICK J. HARRISON, Senior Research Scientist, Office of the Vice President for Research and Compliance, Howard University, Washington, DC

ROMANA HASNAIN-WYNIA, Director, Center for Healthcare Equity and Associate Professor, Research, Feinberg School of Medicine, Northwestern University, Chicago, IL

NINEZ PONCE, Associate Professor, Department of Health Services, UCLA School of Public Health, Los Angeles, CA

WAYNE S. RAWLINS, National Medical Director, Aetna Government Health Plans, Aetna, Hartford, CT

PAUL M. SCHYVE, Senior Vice President, The Joint Commission, Oakbrook Terrace, IL

ALAN M. ZASLAVSKY, Professor of Health Care Policy (Statistics), Harvard Medical School, Boston, MA

Study Staff

Cheryl Ulmer, Project Director

Bernadette McFadden, Research Associate

Michelle Bruno, Research Associate

Adam Schickedanz, Mirzayan Science and Technology Fellow¹

Cassandra Cacace, Senior Program Assistant

Roger Herdman, Board Director, Board on Health Care Services

¹ Served through May 2009.

Reviewers

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the NRC's Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We wish to thank the following individuals for their review of this report:

OLIVIA CARTER-POKRAS, Department of Epidemiology and Biostatistics, College of Health and Human Performance, University of Maryland, College Park, MD
SIMON P. COHN, Kaiser Permanente Medical Care Program, Oakland, CA
HAROLD P. FREEMAN, National Cancer Institute, Rockville, MD and Columbia University, New York, NY
DEANNA L. JANG, Asian & Pacific Islander American Health Forum, Washington, DC
JENNIE R. JOE, Native American Research and Training Center, University of Arizona, Tucson, AZ
ERIC B. LARSON, Center for Health Studies, Group Health Cooperative of Puget Sound, University of Washington, Seattle, WA
DENISE LOVE, National Association of Health Data Organizations, Salt Lake City, UT
JOHN LUMPKIN, Health Care Group, Robert Wood Johnson Foundation, Princeton, NJ
MARY A. PITTMAN, Public Health Institute, Oakland, CA
KENNETH PREWITT, School of International and Public Affairs, Columbia University, New York, NY

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations nor did they see the final draft of the report before its release. The review of this report was overseen by Faith Mitchell, Ph.D, Grantmakers In Health, and Edward B. Perrin, University of Washington. Appointed by the National Research Council and Institute of Medicine, they were responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests entirely with the authoring committee and the institution.

Foreword

The Institute of Medicine report, *Unequal Treatment: Confronting Racial and Ethnic Disparities in Healthcare* (2002) called attention to poorer access to health care and worse health outcomes among certain racial and ethnic groups. According to reports from the Agency for Healthcare Research and Quality and others, disparities in the quality of care and in health outcomes persist. Accelerating progress toward eliminating these disparities depends in part on our ability to identify and track experiences in health care among individuals from a variety of racial and ethnic backgrounds and who speak a variety of languages other than English.

This report offers an approach to identifying racial, ethnic, and language categories that bear on disparities in health care and health outcomes. Extending beyond the broad racial and ethnic categories used by the Office of Management and Budget, this report provides a more granular classification of ethnicity and language needs. This kind of standardized approach to classification will both help measure progress in eliminating disparities and assure that comparisons across different settings are based on similar groupings of individuals.

I want to express my appreciation to the subcommittee and staff for the tremendous effort that has gone into this report. Their work represents another positive step toward the goal of high quality health care for everyone.

Harvey V. Fineberg, M.D., Ph.D.
President, Institute of Medicine
August 2009

Preface

Calling attention to the need for improvement in quality of care has been a central theme for many Institute of Medicine (IOM) reports. *Crossing the Quality Chasm: A New Health System for the 21st Century* noted significant shortcomings in the nation's health care delivery system in terms of safety, effectiveness, timeliness, efficiency, patient-centeredness, and equity while *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health-care* documented that in a variety of organizational settings and clinical domains, members of racial and ethnic minority groups receive poorer quality care than their White counterparts.

While many studies published since the 2003 release of *Unequal Treatment* have shown similar patterns, there is evidence of some progress. Disparities in some domains (e.g., process of care measures such as use of beta blockers or aspirin after heart attack) have been shown to be shrinking over time for some populations but not others. Individual health plans, hospitals, and medical groups have organized quality improvement projects aimed at reducing disparities and have succeeded in doing so. The underlying reasons for disparities are increasingly understood so that initiatives to address disparities can be focused on factors that are likely to have the greatest positive effect. The impact of language, culture, and socioeconomic status, along with race and ethnicity, are also more clearly understood. Yet studies reveal that disparities remain on both process of care and outcome measures.

Continued work in addressing disparities requires the collection and use of data on race, ethnicity, and language in all health and health care data systems, as called for in 2004 by the National Research Council report, *Eliminating Health Disparities: Measurement and Data Needs*. These data provide the opportunity to monitor and analyze disparities, and are informative in identifying individuals and groups to whom quality improvement or other interventions can be directed. Across a range of organizational levels, from the AHRQ *National Health-care Disparities Report* at one end, to the work carried out by individual physician offices and community health centers at the other, the collection and use of data on race, ethnicity, and language are key parts of the process of identifying health care needs and eliminating disparities.

Quality improvement can be organized as a collaborative effort at a local, regional, statewide, or even national level. Even when projects are carried out by individual organizations, the process of benchmarking involves sharing of information from organization to organization. For some quality improvement projects, literal data sharing is important, as an entity collecting race, ethnicity, or language data (e.g., a multispecialty group practice) may provide that information to another entity (e.g., a managed care plan) in order for the second entity to use the information for analyses of quality of care data. Additionally, regional, state, and national health care agencies may wish to pool data from individual organizations to address disparities in a broader geographic context.

The collection of data on race, ethnicity, and language will, in principle, have the greatest impact if it is done according to standards that allow for comparison of data across organizations, sharing of individual-level data from one to another, and combining of data from multiple sources. The Subcommittee on Standardized Collection of Race/Ethnicity Data for Healthcare Quality Improvement was asked to examine the issue of how data on race, ethnicity, and language are collected in various contexts associated with health care, and to offer recommendations on standardization of the categories for these variables. This report addresses data collection challenges and proposes a framework for moving forward with standardized data collection across health care entities. Previous reports have reiterated the importance of collecting more detailed ethnicity data than are captured by the OMB standard categories; this report proposes templates of granular ethnicity and language categories for national adoption so that entities wishing to collect detailed data can do so in systematic, uniform ways. The recommendations presented here provide guidance to entities on data collection to support their efforts to improve quality and eliminate disparities.

David R. Nerenz, Ph.D., *Chair*

Acknowledgments

The subcommittee and staff are grateful to many individuals and organizations who contributed to this study. Most specifically, the subcommittee would first like to thank the IOM Committee on Future Directions for the National Healthcare Quality and Disparities Reports for their guidance and comments on the report. The committee members include:¹

SHEILA BURKE (*Chair*), Faculty Research Fellow, Malcolm Weiner Center for Social Policy, John F. Kennedy School of Government, Harvard University, Boston, MA

ANNE BEAL, Assistant Vice President, The Commonwealth Fund, New York, NY

E. RICHARD BROWN, Professor, UCLA School of Public Health and Director, UCLA Center for Health Policy Research, Los Angeles, CA

MARSHALL H. CHIN, Professor of Medicine, University of Chicago, Chicago, IL

JOSE J. ESCARCE, Professor of Medicine, Division of General Internal Medicine and Health Services Research, UCLA School of Medicine, Los Angeles, CA

KEVIN FISCELLA, Associate Professor, University of Rochester, Rochester, NY

ELLIOT S. FISHER, Professor of Medicine and Community and Family Medicine, Dartmouth Medical School, and Director, Center for Health Policy Research, Dartmouth Institute for Health Care Policy and Clinical Practice, Lebanon, NH

DAWM M. FITZGERALD, CEO, QSource, Memphis TN

FOSTER GESTEN, Medical Director, Office of Health Insurance Programs, New York State Department of Health, Albany, NY

BRENT C. JAMES, Chief Quality Officer and Executive Director, Intermountain Health Care, Inc. Institute for Health Care Delivery Research, Salt Lake City, UT

JEFFREY KANG, Chief Medical Officer and Senior Vice President for Medical Strategy and Policy, CIGNA Corporation, Hartford, CT

SHARON-LISE T. NORMAND, Professor, Department of Health Care Policy, Harvard Medical School, Boston, MA

CHRISTOPHER QUERAM, President/CEO, Wisconsin Collaborative for Healthcare Quality, Middleton, WI

¹ Subcommittee Members, Ignatius Bau, David Nerenz, and Paul Schyve are also members of the Committee.

SARAH SCHOLLE, Assistant Vice President for Research and Analysis, National Committee for Quality Assurance, Washington, DC

BRUCE SIEGEL, Director, Center for Health Care Quality, The George Washington University School of Public Health and Health Services, Washington, DC

The subcommittee acknowledges Constance Citro, director of the Committee on National Statistics, National Research Council of The National Academies, for her insight and contacts within the statistical community as well as the help of Thomas Plewes in locating materials on limited English proficiency.

In addition, the subcommittee benefited from the testimony before the committee and subcommittee during public workshops: Karen Adams (National Quality Forum), Donald Berwick (Institute for Healthcare Improvement), Andrew Bindman (UCSF and CA Medicaid Research Institute), Carolyn Clancy (Agency for Healthcare Research and Quality), Kathryn L. Coltin (Harvard Pilgrim Health Care), Brenda Edwards (Division of Cancer Control & Population Sciences, National Cancer Institute), Marc Elliott (RAND Corporation), Irene Fraser (Agency for Healthcare Research and Quality), Allen Fremont (RAND Corporation), Ron Hays (Division of General Internal Medicine and Health Services Research, UCLA), Karen Humes (U.S. Census Bureau), Deeana Jang (Asian & Pacific Islander American Health Forum, speaking on behalf of Out of Many, One's Health Data Taskforce), Marjorie Kagawa-Singer (UCLA School of Public Health), Karen Kmetik (American Medical Association and The Physician Consortium for Performance Improvement), David Lansky (Pacific Business Group on Health), Nicole Lurie (RAND Corporation, Center for Population Health and Health Disparities), Jennifer Madans (National Center for Health Statistics), Paul McGann (Office of Clinical Standards and Quality, Centers for Medicare and Medicaid Services), Ernest Moy (Agency for Healthcare Research and Quality), Marsha Regenstein (The George Washington University), Thomas Reilly (Office of Research, Development and Information, Centers for Medicare and Medicaid Services), Michael Rodriguez (Department of Family, Medicine, David Geffen School of Medicine, UCLA), Patrick Romano (Divisions of General Medicine and General Pediatrics, Center for Healthcare Policy and Research, UC Davis), Joachim Roski (Engelberg Center for Health Care Reform, The Brookings Institution), Maribeth Shannon (Market and Policy Monitor Program, California HealthCare Foundation), Gayle Tang (National Diversity, Kaiser Permanente), Kalahn Taylor-Clark (Engelberg Center for Healthcare Reform, The Brookings Institution), Grace Ting (Health Equities Programs, Wellpoint, Inc.), Katherine K. Wallman (U.S. Office of Management and Budget), Thomas Williams (Integrated Healthcare Association), and Mara Youdelman (National Health Law Program).

Many others provided valuable advice on the issues under study; these include: Mona L. Bormet (Asian & Pacific Islander American Health Forum), Erin Bowman (California Health Care Safety Net Institute), Rita Carreón (America's Health Insurance Plans), Olivia Carter-Pokras (University of Maryland), Coralie Chan (Kaiser Permanente), Francis Frasier (Indian Health Service), Sundak Ganesan (Centers for Disease Control and Prevention Vocabulary and Messaging Team), Sheldon Greenfield (University of California, Irvine), Kirk Greenway (Indian Health Service), Brady Hamilton (National Center for Health Statistics), William E. Hammond (Duke University), George Isham (HealthPartners), Wendy Jameson (California Health Care Safety Net Institute), Charles Jarvis (NextGen, Executive Team HIMSS EHRA), Ashish Jha (Harvard School of Public Health), Sherrie Kaplan (University of California, Irvine), Theodore Klein (Klein Consulting), Diane Louise Leach (Indian Health Service), Mark Leavitt (Certification Commission for Healthcare Information Technology), Denise Love (National Association of Health Data Organizations), Martin Martinez (California Pan-Ethnic Health Network), Vickie M. Mays (Department of Health Services, UCLA), Mark McClellan (The Brookings Institution), JeanHee Moon (Center for Health Care Strategies), Jeannette Noltenius (Out of Many, One), Edna Paisano (Indian Health Service), Ruth Perot (Summit Health Institute for Research and Education), Daniel Pollack (Centers for Disease Control and Prevention), Kenneth Prewitt (School of International and Public Affairs, Columbia University), Alisa Ray (Certification Commission for Healthcare Information Technology), C. Sue Reber (Certification Commission for Healthcare Information Technology), Richard Pride (University of Mississippi Medical Center), Bob Rehm (America's Health Insurance Plans), Yvette Roubideaux (University of Arizona College of Medicine), Barbara Rudolph (The Leapfrog Group), Hyon Shin (U.S. Census Bureau), Jane Sisk (National Center for Health Statistics, Centers for Disease Control and Prevention), Benjamin P. Smith (Indian Health Service), Brian Smedley (Joint Center for Political and Economic Studies), Phillip L. Smith (Indian Health Service), Benjamin Steffen (Maryland Health Care Commission), Otilia

Tiutin (Contra Costa Health Plan), Alan Trachtenberg (Indian Health Service), William Vega (David Geffen School of Medicine, UCLA), Lucie Vogel (Indian Health Service), Robin Weinick (Institute of Health Policy, Massachusetts General Hospital), and Ellen Wu (California Pan-Ethnic Health Network).

California state government representatives were generous in their time discussing issues with respect to collection of data and implementation of SB 853. These included: Cindy Ehnes and Hattie Hanley of the Department of Managed Health Care; Shelley Rouillard and Ernesto Sanchez, Managed Risk Medical Insurance Board; David Carlisle, Candace Diamond, Serena Beltran, and Ron Spingarn, Office of Statewide Health Planning and Development; Sandra Perez and Ed Mendoza, California Office of the Patient Advocate; and Rita Marowitz, Medi-Cal Managed Care Division.

Many within the Institute of Medicine (IOM) were helpful throughout the study process, including Rose Martinez, Lyla Hernandez, Karen Anderson, Sharyl Nass, and Laura Levitt. In addition, the staff would like to thank Clyde Behney, Lauren Tobias, Abbey Meltzer, Jordan Wyndelts, Jackie Turner, Vilija Teel, Christine Stencel, and Linda Kilroy for their continuing support throughout the project to ensure release of this expedited report. The committee also wishes to acknowledge the editing provided by Rona Briere.

Funding for this study was provided by the Agency for Healthcare Research and Quality (AHRQ). The subcommittee appreciates its support for the project as well as substantive support from AHRQ staff, particularly Carolyn Clancy, Ernest Moy, and Roxanne Andrews. The California Endowment has provided additional funding to ensure widespread distribution of this report's summary brief in Spanish and Chinese.

Contents

SUMMARY	1
1 INTRODUCTION	13
2 EVIDENCE OF DISPARITIES AMONG ETHNICITY GROUPS	31
3 DEFINING CATEGORIZATION NEEDS FOR RACE AND ETHNICITY DATA	61
4 DEFINING LANGUAGE NEED AND CATEGORIES FOR COLLECTION	93
5 IMPROVING DATA COLLECTION ACROSS THE HEALTH CARE SYSTEM	127
6 IMPLEMENTATION	147
APPENDIXES*	
A Acronyms and Abbreviations	163
B Legislation Used in Report	167
C Workshop Agendas	171
D Subcommittee Member and Staff Biographies	177
E SUBCOMMITTEE TEMPLATE: Comparison of Granular Ethnicity Category Lists and Rollup Schemes	
F Granular Ethnicities with No Determinate OMB Race Classification	
G Kaiser Permanente: Evolution of Data Collection on Race, Ethnicity, and Language Preference Information	
H Contra Costa Health Plan Language Assistance Database and Ethnicity Categories	
I SUBCOMMITTEE TEMPLATE: Comparison of Spoken Language Categories and Coding	

*Appendixes E through I can be found online at <http://www.iom.edu/datastandardization>.

Boxes, Figures, and Tables

Summary

Box

- S-1 Statement of Task: Subcommittee on Standardized Collection of Race/Ethnicity Data for Healthcare Quality Improvement, 2

Figure

- S-1 Recommended variables for standardized collection of race, ethnicity, and language need, 3

Chapter 1

Boxes

- 1-1 Barriers to Collection of Race, Ethnicity, and Language Data, 18
1-2 The 1997 OMB Revisions to the Standards for the Classification of Federal Data on Race and Hispanic Ethnicity, 20
1-3 Statement of Task: Subcommittee on Standardized Collection of Race/Ethnicity Data for Healthcare Quality Improvement, 23

Figures

- 1-1 A framework for reducing disparities in health care systems, 15
1-2 Williams, Lavizzo-Mourey, and Warren's framework for understanding the relationships between race, medical/health care, and health, 18
1-3 Overview of purposes and uses of race, ethnicity, and language data to guide subcommittee's investigation of issues of categorization and collection, 24

Table

- 1-1 Categories and Definitions Promulgated by the OMB and the U.S. Bureau of the Census, 17

Chapter 2

Tables

- 2-1 Census 2000 Population by Race and Hispanic Ethnicity, 33
- 2-2 Selected Characteristics of the Hispanic/Latino/Spanish Population in the United States, 37
- 2-3 Selected Characteristics of the Black Population in the United States, 42
- 2-4 Selected Characteristics of the Asian Population in the United States, 45
- 2-5 Selected Characteristics of the NHOPI Population in the United States, 49
- 2-6 Selected Characteristics of the White Population in the United States, 51
- 2-7 Selected Characteristics of the American Indian or Alaska Native (AIAN) Population in the United States, 54

Chapter 3

Boxes

- 3-1 Race and Ethnicity Categories in the HCUP Databases, 62
- 3-2 The Use of Data Linkages to Improve Data Coverage and Quality in Cancer Registries, 69
- 3-3 The Challenge of Categorizing Filipino Respondents, 71
- 3-4 Realizing the Necessity of Collecting Data: The University of Mississippi Medical Center, 80

Figures

- 3-1 Reproduction of questions of race and Hispanic origin from Census 2000, 71
- 3-2 Geographic distribution of the Asian population, 76
- 3-3 CDC ethnicities rolled up to the OMB minimum categories for race and Hispanic ethnicity with subcommittee annotations, 82
- 3-4 Models for data collection instruments to collect race, Hispanic ethnicity, and granular ethnicity data, 88

Tables

- 3-1 OMB Race and Hispanic Ethnicity Categories According to a One- and Two-Question Format, 62
- 3-2 Race and Ethnicity Categories Collected by Various Data Sources, 64
- 3-3 Race and Hispanic Ethnicity Categories Used by State Medicaid and CHIP Programs, 67
- 3-4 Hispanic and Not-Hispanic Population Distribution by Race for the United States: 2000, 72
- 3-5 Comparison of Granular Rthnicity Categorization and Coding Systems, 78
- 3-6 Examples of Instructions, Phrasing, and Terminology to Capture Race and Ethnicity Data, 86

Chapter 4

Boxes

- 4-1 Language Concordance Between Patients and Providers, 95
- 4-2 Assessing Whether Language Assistance Needs are Met, 96

Figures

- 4-1 Census 2000 questions about language, 100
- 4-2 Karliner algorithm, 104
- 4-3 Most spoken languages in North Dakota, Minnesota, Texas and Maine, 2005, 112
- 4-4 Number of languages spoken in each state, 114

Tables

- 4-1 Summary of Question Types and Categories, 103
- 4-2 Correlations Between Self-Reported English Ability in Speaking, Reading, and Writing, 106
- 4-3 Relationship of Speaking and Reading Ability, 107
- 4-4 Language Categories in Selected Collection Instruments, 110

Chapter 5

Boxes

- 5-1 Statewide Race and Ethnicity Data Collection: Massachusetts, 131
- 5-2 Collecting and Using Data: The Alliance of Chicago Community Health Services, 132
- 5-3 Collecting Data in Small Physician Practices, 133
- 5-4 Successful Collection of Data by a Health Plan: Aetna, 134
- 5-5 Standardizing Direct Data Collection, 138
- 5-6 The Use of Indirectly Collected Data by a Health Plan: Wellpoint, Inc., 142

Figures

- 5-1 A snapshot of data flow in a complex health care system, 129
- 5-2 Opportunities to collect data within the health care system, 130

Chapter 6

Figure

- 6-1 Recommended variables for standardized collection of race, ethnicity, and language need, 149

