HSR Health Services Research

DOI: 10.1111/j.1475-6773.2010.01143.x
SPECIAL ISSUE: HEALTH SERVICES RESEARCH IN 2020

Data and Measurement Issues in the Analysis of Health Disparities

Linda T. Bilheimer and Richard J. Klein

Objective. To describe measurement challenges and strategies in identifying and analyzing health disparities and inequities.

Methods. We discuss the limitations of existing data sources for measuring health disparities and inequities, describe current strategies to address those limitations, and explore the potential of emerging strategies.

Principal Findings. Larger national sample sizes are necessary to identify disparities for major population subgroups. Collecting self-reported race and granular ethnicity data may reduce some measurement errors, but it raises other methodological questions. The assessment of health inequities presents particular challenges, requiring analysis of the interactive effects of multiple determinants of health. Indirect estimation and modeling methods are likely to be important tools for estimating health disparities and inequities for the foreseeable future.

Conclusions. Interdisciplinary training and collaborative research models will be essential for future disparities research. Evaluation of evolving methodologies for assessing health disparities should be a priority for health services researchers in the next decade.

Key Words. Disparity, inequity, measurement, surveys, indirect estimation

OVERVIEW

Disparities in health status, health care access, and health care outcomes among subgroups of the population emerged as a major public policy concern 25 years ago. The 1985 *Report of the Secretary's Task Force on Black and Minority Health* called for multiple policy responses to excess deaths among black and other racial and ethnic subgroups (U.S. Department of Health and Human Services 1985; Centers for Disease Control and Prevention 1986). The report emphasized the need for greater coordination among federal agencies to address disparities, better data on the health of racial and ethnic subgroups, and more research to investigate the factors affecting their health. Subsequent federal policy actions included the establishment of the Office of Minority Health and the Initiative to Eliminate Racial and Ethnic Disparities in Health (Satcher 1999).

The Healthcare Research and Quality Act of 1999 (P.L. 106–29) broadened the scope of concern to incorporate disparities based on socioeconomic status, race, and ethnicity, as well as disparities among other priority populations, including women, children, elderly people, people with special health care needs, and those living in rural areas and inner cities. The Act established the Agency for Healthcare Research and Quality (AHRQ), directing it to produce an annual National Healthcare Disparities Report (NHDR). Shortly thereafter, the Administration released the Healthy People 2010 (HP2010) goals for the Nation, encompassing two distinct goals: to increase the length and quality of life of the population, and to eliminate health disparities associated with race and ethnicity, socioeconomic status, gender, age, urbanicity, disability status, and sexual orientation (U.S. Department of Health and Human Services 2000). The initiative called for improved data collection to monitor changes in social and economic disparities. Despite this focus, health disparities remained unchanged for most HP2010 objectives during the first half of the decade (U.S. Department of Health and Human Services 2006).

Recently, the concept of health inequity, connoting injustice, has become an important part of the assessment of health disparities (Carter-Pokras and Baquet 2002; Braveman and Gruskin 2003; Braveman 2006). HP2010 called for a multidisciplinary approach to achieving health equity, involving improvements in health, education, housing, labor, justice, transportation, agriculture, and the environment (U.S. Department of Health and Human Services 2000). The first NHDR also distinguished between disparity and inequity, concluding that not all differences observed at the individual level necessarily reflect inequity (U.S. Department of Health and Human Services 2003). Achieving health equity is a central focus of Healthy People 2020 (HP2020), with an emphasis on social, economic, and environmental determinants of health (U.S. Department of Health and Human Services 2008).

Since the initiation of policies to address health disparities in the 1980s, data and measurement problems have presented perennial challenges for

---

Address correspondence to Linda T. Bilheimer, Ph.D., Office of Analysis and Epidemiology, Centers for Disease Control and Prevention, National Center for Health Statistics, 3311 Toledo Rd, Hyattsville, MD 20782; e-mail: lbilheimer@cdc.gov. Richard J. Klein, M.P.H., is with the Health Promotion Statistics Branch, Office of Analysis and Epidemiology, Centers for Disease Control and Prevention, National Center for Health Statistics, Hyattsville, MD.

researchers and policy makers (Ver Ploeg and Perrin 2004). Most attention has focused on data limitations relating to race and ethnicity, but problems also arise in identifying disparities in other priority subpopulations. The recently enacted Patient Protection and Affordable Care Act (P.L. 111–48) therefore requires any federally conducted or supported health care or public health program, activity, or survey to collect and report data on race, ethnicity, sex, primary language, and disability status, to the extent feasible; to collect these data at the smallest geographic level possible; and to have sufficient data to generate reliable estimates for subgroups of these populations.

In this paper, we discuss challenges in identifying and analyzing both health disparities and inequities, and we explore emerging ideas about future measurement strategies. The wide range of approaches and disciplines that these strategies encompass highlights the importance of multidisciplinary approaches to disparities analysis, with important implications for training researchers (Carey and Howard 2007).

IDENTIFYING DISPARITIES: CHALLENGES AND STRATEGIES

Identifying health disparities requires data on health status and individual determinants of health for subgroups of the population, at the national, state, and community levels—data that are frequently difficult to obtain. Problems include small sample sizes, missing data, and measurement errors.

Sample Size Limitations

Small sample sizes restrict the ability of researchers to measure disparities, even for some of the major racial, ethnic, and socioeconomic subgroups at the national level (Bilheimer and Sisk 2008). The difficulties are compounded at the state and community levels, where actions to address disparities are most likely to occur (Gold, Dodd, and Neuman 2008). Only a few states have their own household health surveys in addition to the Behavioral Risk Factor Surveillance System (BRFSS), which can identify a limited number of racial, ethnic, and socioeconomic groups, but often cannot provide substate estimates for subgroups of the population. For the few surveys that collect detailed racial and ethnic information, within-group variation—among Asian, Hispanic, or American Indian and Alaska Native subgroups, for example—is also difficult to identify because of sample size limitations. Available studies reveal wide variations among such subgroups, demonstrating the importance of more detailed analyses (see, e.g., Barnes, Adams, and Powell-Griner 2008, Brown 2008). Possible data collection and analysis strategies to address small sample size problems include the following.

Oversampling. National surveys routinely oversample populations of policy interest. For disparity populations in particular, the National Health Interview Survey (NHIS) oversamples blacks, Hispanics, and Asians. The National Health and Nutrition Examination Survey (NHANES) currently oversamples non-Hispanic blacks, Hispanics, and low-income people, and it will begin oversampling Asian Americans in 2011. The Medical Expenditure Panel Survey (MEPS) oversamples low-income people, based on predicted poverty status, in addition to the racial/ethnic oversampling that automatically occurs by virtue of its ties to the NHIS sample (Wun, Ezzati-Rice, and Baskin 2008). At the state level, the California Health Interview Survey oversamples Vietnamese and Koreans, and received special funding in 2001 to oversample American Indians and Alaska Natives (UCLA Center for Health Policy Research).

Pooled Data. Combining survey data from several years is a common strategy to address small sample size problems. In assessing racial disparities in hospitalizations for lower extremity amputations among people with diabetes, for example, the NHDR compared rates for the period 2001–2003 to rates for 2004–2006 (U.S. Department of Health and Human Services 2009). This approach requires data collection instruments to remain consistent over time and—depending on the number of years that must be combined—may limit the ability to track short-term responses to policy changes.

Targeted Periodic Surveys. Surveys such as the Community Health and Nutrition Examination Surveys (CHANES) provide an alternative approach for studying disparities in small population subgroups. Like NHANES, CHANES integrates data from direct physical examinations, and clinical and laboratory tests with personal interviews on health and nutrition status. The "community" can be a racial or ethnic group, as exemplified by the Hispanic HANES (HHANES) initiative, or groups defined by geographic location, socioeconomic status, or other factors (see, e.g., Marwick 1991, New York City Department of Health and Mental Hygiene). HHANES, conducted in 1982–1984, provided extensive insights into the health of Hispanics and disparities among Hispanic subgroups (although the data were not nationally representative). Similar studies among the Asian, American Indian or Alaska Native, and Native Hawaiian and Other Pacific Islander populations could prove equally informative, as might a repeat of HHANES. Because of the costs, such targeted surveys probably could not be conducted frequently, but they could provide detailed periodic snapshots of the health status of priority populations.

Multiple Frame Sample Designs. A recent study raised the possibility of using this strategy to facilitate oversampling small populations (Elliott et al. 2008a). Focusing on the American Indian and Alaska Native and Chinese populations, the authors explored several methods for augmenting sample sizes in NHIS, and they identified three potentially promising approaches: complete sampling within households that contain members of a target racial or ethnic group; oversampling selected macrogeographic units; and oversampling from incomplete sampling frames derived from surname and telephone lists. Important considerations include the tradeoffs between the design effects, which reduce the effective sample size, the gains for particular target populations, and costs. In this study, the effectiveness of the three approaches varied between the two population groups. Complete sampling within households showed promise for both groups, as did macrogeographic sampling. But due to their greater residential concentration, the latter approach would be less costly for American Indians and Alaska Natives than for the Chinese population. The incomplete frame approach did not work well for American Indians and Alaska Natives because of the low rate of landline telephone coverage and the lack of a viable surname list. In addition, the cost-effectiveness of all three strategies declined as the size of the supplemental sample increased relative to the base sample.

Modeling Approaches. Researchers are also developing models that use Bayesian statistical methods to produce more precise estimates for subpopulations that are underrepresented in national sample surveys (Elliott et al. 2009). Such approaches "borrow strength" from larger population groups or across time periods to reduce standard errors. Using a modified Kalman filter approach, Elliott and colleagues demonstrated the feasibility of using several years of cross-sectional data to reduce significantly the error of current year estimates for small racial and ethnic groups. These methods raise questions about the use of model-based estimates in policy making and their acceptability to the policy community. Questions might arise if, for example, a modeling approach increased the precision of an estimate for a small group but also changed the underlying point estimate, thereby changing disparity estimates as well.

Mixed Mode Surveys. Given the high cost of obtaining even national-level data on small subpopulations through in-person interviews, experts are now considering alternative approaches to obtain consistent state- and community-level data for subpopulations that combine survey modes. A large national in-person interview survey, such as an expanded NHIS, could produce consistent state estimates for racial and ethnic subpopulations, and serve as an umbrella for state-specific telephone surveys that could be calibrated to the national estimates (Madans 2009).

Missing Data

In many datasets—both surveys and administrative data systems—information on key variables for identifying disparities may be missing or not collected. Income, for example, is often missing or not reported exactly in widely used surveys such as NHIS, and it is not reported at all in several of the most important data systems used in HP2010, such as the National Vital Statistics System, the National Hospital Discharge Survey (NHDS), the National Ambulatory Medical Care Survey (NAMCS), the National Hospital Ambulatory Medical Care Survey (NHAMCS), the Health Care Cost and Utilization Project, and the Youth Risk Behavior Surveillance System. Likewise patient race and ethnicity data are often missing in surveys of health care providers, such as NHDS, NAMCS, and NHAMCS, which collect information from patient records (Sonnenfeld and Sisk 2008). Strategies to address missing data may include imputation, data linkage, or the use of proxy variables. Advances in the use of electronic health records (EHRs) also hold promise for increasing the availability and quality of patient race and ethnicity data. Other possibilities include incentives for providers to collect race and ethnicity information, and regulatory requirements. In the interim, some health plans and health care providers are using indirect estimation techniques to identify racial and ethnic disparities in health care access and quality.

Imputation. Major national surveys routinely use imputation to address missing value problems. NHIS has adopted a multiple imputation approach to handle missing income information that allows for assessment of the variability due to imputation. That imputation appears to correct for biases that occur without imputation and usually results in lower standard errors (Schenker et al. 2006). The MEPS also imputes missing income. NAMCS and NHAMCS both impute missing race and ethnicity information (Sonnenfeld and Sisk 2008). A new imputation approach was introduced in 2006 for those surveys, and validation is ongoing.

Data Linkage. If disparity variables are not available in a dataset, microdata linkage may provide a solution for some types of analyses. The linked NHIS/mortality files, for example, may be used to identify income-related disparities in mortality, which cannot be identified from the mortality statistics alone. Data linkage may also be used to improve the imputation of missing values, as MEPS does through its link to NHIS in imputing income (Banthin and Selden 2006).

Proxy Variables. Researchers sometimes use proxies for missing disparity variables, but socioeconomic variables are not necessarily good proxies for each other. Braveman et al. (2001), citing the relatively low correlation between education and income among mothers in different racial and ethnic groups, argued against using education as a proxy for income in analyzing disparities. Other researchers have used aggregate income or education measures (at the zip code, Census tract, or Census block level) to proxy for an individual's income or education (Krieger 1992; Gornick 2003). But because of potential biases and confounding problems, researchers should use caution in substituting aggregate for individual-level variables (Geronimus, Bound, and Neidert 1996).

EHRs. The current policy focus on the development of EHRs presents an opportunity to address the nonreporting and lack of standardization of patient race and ethnicity data in medical records. An effective health information infrastructure would require only one-time collection of self-reported race/ethnicity/language information, which could then be exchanged with other health care organizations under strict privacy protections (Ulmer, McFadden, and Nerenz 2009). The HIT Policy Committee established by the American Recovery and Reinvestment Act of 2009 was charged with making recommendations on standards, implementation specifications, and certification criteria for a nationwide health information technology infrastructure (P.L. 111–5). The recently released notice of proposed rulemaking for establishing "meaningful use" of EHRs envisions a three-stage approach (Federal Register, 2010). Stage 1 objectives, to be accomplished by 2011, would require providers to document patient race, ethnicity, primary language, insurance type, and gender for at least 80 percent of their patients. The race and ethnicity reporting requirements reflect the current Office of Management and Budget (OMB) standards.

Regulations and Incentives for Providers. A recent report from the Institute of Medicine (IOM) recommended a variety of ways to improve the collection of race and ethnicity data by providers (Ulmer, McFadden, and Nerenz 2009). Requirements to collect such data could be included in payment incentive programs for improving health care quality; regulations for federal and state funded health programs; and accreditation requirements and performance standards.

Indirect Estimation. Some health plans seeking to identify racial and ethnic disparities in health care access and quality are using indirect methods of estimation, as direct collection of this information has proven difficult (Higgins and Taylor 2009). Problems range from IT modifications to challenges in determining the best way to collect race, ethnicity, and language information. Indirect methods use different combinations of surname analysis, geocoding, and Bayesian techniques with effectiveness varying by subgroup (Fiscella and Fremont 2006; Elliott et al. 2008b). The addition of Bayesian techniques to surname analysis and geocoding allows estimation of the probability that an individual is in a particular group, and it appears to improve estimates of race and ethnicity relative to other indirect methods. This approach is useful for determining the racial and ethnic composition of a health plan's enrollees and identifying disparities among groups, but it should not be used to identify the race or ethnicity of specific individuals (High-Value Health Care Project 2010).

Measurement Errors

In the analysis of racial and ethnic health disparities, much attention has focused on the measurement errors that arise from different data collection methods. Medical records and administrative data systems use unknown combinations of self-report by patients, information from other family members, and observation by others. Moreover, although self-report is viewed as a gold standard, as the U.S. population becomes more diverse, people from different countries of origin may not self-identify with any of the racial and ethnic categories that surveys and reporting systems include, which are typically based on the OMB categories (Executive Office of the President 1997). More than 40 percent of Hispanics, for example, did not identify with any of the listed racial groups in the 2000 Census, reporting themselves as "some other race" (Grieco and Nassidy 2001). Measurement errors also result from cross-cultural biases in survey instruments that may disguise health disparities or contribute to apparent differences (see, e.g., Carle 2009, Hillenmaier et al. 2007, Leung et al. 2007). To address these measurement problems, interest is growing in the collection of more granular race and ethnicity data and in cross-cultural measurement research.

Self-Reported Race and Granular Ethnicity Data. Some experts have suggested that both surveys and administrative data systems should move beyond self-report from a limited list of race/ethnicity options—such as the OMB categories provide—and ask people to report their self-identified race and/or ethnicity (Prewitt 2005; Hasnain-Wynia and Baker 2006). This approach raises questions of how to address the potential heterogeneity of responses as well as obtain the denominator data to calculate the rates from which disparities are determined. Proponents maintain that developing algorithms to "roll up" self-identified race/ethnicity into the OMB standard categories is feasible.

The recent IOM report endorsed the idea of rolling up granular ethnicities into the OMB standards, but it did not advocate a completely open-ended approach (Ulmer, McFadden, and Nerenz 2009). Rather, it proposed using standard national lists for both ethnicity and preferred language to assess and report on disparities in health care quality. Organizations would select appropriate categories from those lists to meet their local needs. For people who did not identify with any of the OMB race or Hispanic origin categories, the report proposed rolling up from granular ethnicities to those categories. Because all ethnicities would not roll up to a specific OMB racial category, the report suggested criteria for classifying individuals as having "no determinate OMB classification." If the rolling up approach is widely adopted, a key research question is whether the resulting aggregations produce more or less within-group variation than self-report based on the OMB categories.

Granular Data Collection by Health Plans and Providers. Kaiser Permanente and the state of Massachusetts are among the pioneers in developing standardized approaches to the collection of patient race and ethnicity information. Kaiser's Member Demographic Data Collection Initiative collects self-identified information on race, ethnicity, language, and country of birth, using a modified form of the combined OMB categories and an extensive list of granular ethnicities (Tang 2009). Massachusetts has also developed a standardized race and ethnicity reporting system for hospital discharges based on the OMB categories that incorporates a detailed list of granular ethnicities (Parlato 2009). Their experiences demonstrate the complexities of implementing such changes, the importance of provider training and effective communication with patients, and the length of time necessary to achieve high-quality reporting. In recognition of these complexities, the IOM report recommended the use of indirect methods of estimation until primary data collection efforts improve.

Granular Data Collection in Surveys. Providing more structure to the collection of granular data addresses some problems of response heterogeneity, but the feasibility of this approach for national household surveys is still questionable. Cost constraints limit sample sizes, the reliability of detailed information is uncertain if respondents must select from a large number of possible options, and confidentiality protections limit the degree to which highly granular information can be released. Granularity may also complicate oversampling strategies. Given the concentration of relatively rare racial, ethnic, and linguistic groups in certain states and localities, however, statewide surveys of local communities might allow collection of granular data for specific local populations.

Denominator Issues. Obtaining denominators to calculate rates based on granular data should be straightforward for health care organizations that are focusing on analyzing disparities and inequities among their patient populations; if they succeed in collecting detailed race and ethnicity information on all their patients, they have the denominators. The situation is more complex, however, if they want to assess disparities in health care access in their communities, as they then need the same granular information for the local population as a whole. The Census Bureau now collects race and ethnicity information in both the Census and on the American Community Survey (ACS)—a large continuous demographic survey—but ancestry information is collected only in the ACS. Depending on the population sizes of geographic areas, the ACS releases 1-, 3-, and 5-year estimates, with the first 5-year estimates to be released in 2010 (U.S. Bureau of the Census 2008). One-, 3-, and 5-year estimates will be released annually hereafter. Even combining 5 years of data, however, standard errors of estimates may be large for small racial, ethnic, and ancestry groups. Five-year estimates, moreover, will be slow to pick up rapid changes in population composition. Community-level rates might, therefore, have to be based on aggregated data.

Cross-Cultural Measurement. Cultural variability in responses to survey questions may primarily reflect measurement biases rather than substantive differences in the variable of interest. Recognition of this problem has led to calls for more studies to assess measurement equivalence, which requires expertise in both quantitative and qualitative research methods (Ramirez et al. 2005; McHorney and Fleishman 2006). Issues of concern include a lack of consensus on the conceptual meaning of equivalence, whether the emphasis of measurement should be on shared methods or shared meaning, and the effects of cultural value systems on survey responses (Johnson 2006).

IDENTIFYING INEQUITIES: A MULTIDIMENSIONAL PROBLEM

To assess inequities, analysts must determine whether a health or health care disparity is due to differences in social, economic, environmental, or health care resources, which requires analysis of the interactive effects of multiple determinants of health, including individual characteristics. Data needs are more complex than those for identifying disparities and include information on such factors as the health care system, housing, transportation, education, and the environment, many of which operate at the community level. Incorporating multiple domains into a single survey is difficult, and surveys are not always the appropriate tools for collecting information on community—as opposed to individual—determinants of health. Moreover, not all of the potentially important factors—the effects of past and present discrimination on racial and ethnic health disparities, for example—are readily measurable. The breadth of the factors that must be considered may take health services researchers out of their usual domains and require collaborations with researchers in other disciplines. These collaborations could determine which factors should be considered when investigating the causes of health disparities, and how to operationalize and measure them. A variety of indirect estimation techniques and linkage strategies are emerging as potential tools to identify some of the determinants of disparities.

Small Area Estimation

Researchers have used strategies that combine data from household surveys, such as the BRFSS, with data from the Census and other sources to produce small area estimates of health disparities and their determinants, including community factors (see, e.g., Li et al. 2009a, b, Schneider et al. 2009). The shift from the Census long form to the ACS, which is much larger than other national household surveys and provides extensive information on many of the social and economic determinants of health—now including health insurance coverage—should help to improve the quality of such estimates (Davern 2008).

Geospatial Techniques

Geocoding and/or linkage to administrative datasets may also be used to address multidimensional data needs, in addition to their use in identifying health care disparities at the health plan level. A recent study, for example, linked the geocoded NHANES III survey to the Census, to explore the association between neighborhood socioeconomic status and consumption of fruits and vegetables (Dubowitz et al. 2008) Other studies have assessed inequities in access to nutritious food, using geocoding and commercially available lists to link restaurant and grocery store information to surveys and the Census (Larson, Story, and Nelson 2009). Researchers have also used geocoding to explore racial disparities at different levels of geographic aggregation as an indirect means to identify health disparities that reflect modifiable versus innate risk factors (Meliker et al. 2009). These studies illustrate the potential of geospatial techniques for the analysis of inequities, as well the types of methodological concerns that may arise, such as questions about the definitions of neighborhoods, the validity of the establishment data in commercial lists, and the accuracy of automated geocoding tools.

DISCUSSION

This brief overview of data issues in the assessment of health disparities and inequities points to several conclusions and also raises some concerns for researchers and policy analysts to consider.

Larger National Sample Sizes Are Necessary to Identify Disparities for Major Population Subgroups at Both the National and State Levels

Sample sizes large enough to produce annual national and state estimates would facilitate national policies to address disparities and provide an infrastructure on which state-specific surveys that produced community-level estimates could build, making local estimates more comparable across the states.

But expanding sample sizes to achieve this goal would be costly, and it would require considerable investment in data systems.

Provider Incentives and Regulations Could Result in Better Self-Reported Patient Race, Ethnicity, and Language Data

The recent IOM report envisioned a variety of incentive and regulatory approaches to improve the reporting of race, granular ethnicity, and language data by providers. As race, ethnicity, language, and gender characteristics are included in the Stage 1 "meaningful use" standards, providers will, indeed, have incentives to collect some information, but not necessarily at the granular level that the IOM subcommittee is seeking (because the meaningful use race and ethnicity criteria are based on the OMB categories).

Collecting Self-Reported Race and Granular Ethnicity Data Has the Potential to Reduce Some Measurement Errors But Raises Other Methodological Questions

As both Kaiser Permanente and Massachusetts have shown, with time, training, and effective communication, health care organizations can collect more self-reported race and granular ethnicity data on their patients. Kaiser Permanente is also collecting detailed language data, and Massachusetts is considering doing so. But collecting granular data in household surveys seems less feasible. At issue are both the validity of rolling up from granular ethnicities to higher levels of aggregation and, ultimately, to the OMB categories, and how to calculate rates based on granular data. Lack of denominator data may require use of the aggregated information to calculate community-level rates, which reinforces the importance of further validity studies.

Indirect Estimation Techniques Are Likely to Continue to Be Important Tools for Assessing Health Disparities and Inequities

Researchers are using indirect methods to produce more precise estimates of disparities in small subpopulations, model community-level and socioeconomic determinants of disparities, and identify racial and ethnic disparities in health care quality at the health plan and provider levels. Over time, EHR improvements should enable more plans and providers to obtain self-reported race and ethnicity information. But given the difficulties experienced to date, indirect estimation will probably be necessary for the foreseeable future. Researchers and policy makers need more discussion about the appropriate use and acceptability of indirect estimation in the policy process.

Interdisciplinary Training and Collaborative Research Models Will Be Essential for Future Disparities Research

Emerging methodologies encompass a wide and growing array of tools from multiple social science and statistical disciplines, requiring researchers who are methodologically adept, with the ability to work in multidisciplinary teams, and posing challenges for traditional graduate programs in health services research. Evaluation of these evolving methodologies should be a priority for researchers in the next decade.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: The findings and conclusions in this paper are those of the authors and do not necessarily represent the views of the Centers of Disease Control and Prevention, National Center for Health Statistics. The authors received no financial or material support for this project. They received helpful comments from Jennifer Madans, Diane Makuc, James Craver, and Marsha Gold. This paper has been cleared for publication through the Centers for Disease Control and Prevention, National Center for Health Statistics clearance process.

Disclosures: None.
Disclaimers: None.

REFERENCES

Banthin, J. S., and T. M. Selden. 2006. *Income Measurement in the Medical Expenditure Panel Survey*. Agency for Health Care Research and Quality Working Paper: No. 06005 [accessed on July 2, 2010]. Available at http://www.meps.ahrq.gov/mepsweb/data%5Ffiles/publications/workingpapers/wp%5F06005.pdf

Barnes, P. M., P. F. Adams, and E. Powell-Griner. 2008. *Health Characteristics of the Asian Adult Population: United States, 2004–2006*. Advance Data from Vital and Health Statistics: No. 394. Hyattsville, MD: National Center for Health Statistics [accessed on July 10, 2009]. Available at http://www.cdc.gov/nchs/data/ad/ad394.pdf

Bilheimer, L., and J. Sisk. 2008. "Collecting Adequate Data on Racial and Ethnic Disparities in Health: The Challenges Continue." *Health Affairs* 27: 383–91.

Braveman, P. 2006. "Health Disparities and Health Equity: Concepts and Measurement." *Annual Review of Public Health* 27: 167–94.

Braveman, P., C. Cubbin, K. Marchi, S. Egerter, and G. Chavez. 2001. "Measuring Socioeconomic Status/Position in Studies of Racial/Ethnic Disparities: Maternal and Infant Health." *Public Health Reports* 116: 449–63.

Braveman, P., and S. Gruskin. 2003. "Defining Equity in Health." *Journal of Epidemiology and Community Health* 57: 254–8.

Brown, E. R. 2008. "The Changing Face of American Communities: "No Data, No Problem." America in Transition: A View from California." Institute of Medicine Roundtable on Health Disparities [accessed on July 2, 2010]. Available at http://www.iom.edu/~/media/Files/Activity%20Files/SelectPops/.../Brown.ashx

Carey, T. S., and D. L. Howard. 2007. "The Development of the Field of Health Disparities Research: The Role of Cross-Disciplinary and Institutional Collaboration." *Harvard Health Policy Review* 8: 136–44.

Carle, A. 2009. "Cross-Cultural Invalidity of Alcohol Dependence Measurement across Hispanics and Caucasians in 2001 and 2002." *Addictive Behaviors* 34: 43–50.

Carter-Pokras, O., and C. Baquet. 2002. "What Is a 'Health Disparity'?" *Public Health Reports* 117: 426–34.

Centers for Disease Control and Prevention. 1986. "Perspectives in Disease Prevention and Health Promotion: Report of the Secretary's Task Force on Black and Minority Health." *Mortality and Morbidity Weekly Report* 35: 109–12.

Davern, M. 2008. *The Impact of the American Community Survey on Small Area Estimation.* Presentation at the 25th Annual Behavioral Risk Factor Surveillance System Conference, Orlando, FL [accessed on July 10, 2009). Available at http://www.shadac.org/files/Pres_BRFSS2008Mar17_Davern.pdf

Dubowitz, T., M. Heron, C. E. Bird, N. Lurie, B. K. Finch, R. Basurta-Davila, H. Hale, and J. J. Escarce. 2008. "Neighborhood Socioeconomic Status and Fruit and Vegetable Intake among Whites, Blacks, and Mexican Americans in the United States." *American Journal of Clinical Nutrition* 87: 1883–91.

Elliott, M., B. Finch, D. Klein, S. Ma, D. Do, M. Beckett, N. Orr, and N. Lurie. 2008a. "Sample Designs for Measuring the Health of Small Racial/Ethnic Subgroups." *Statistics in Medicine* 27: 4016–29.

Elliott, M. N., A. Fremont, P. A. Morrison, P. Pantoja, and N. Lurie. 2008b. "A New Method for Estimating Race/Ethnicity and Associated Disparities Where Administrative Records Lack Self-Reported Race/Ethnicity." *Health Services Research* 43 (5, part I): 1722–36.

Elliott, M. N., D. F. McCaffrey, B. K. Finch, D. J. Klein, N. Orr, M. K. Beckett, and N. Lurie. 2009. "Improving Disparity Estimates for Rare Racial/Ethnic Groups with Trend Estimation and Kalman Filtering: An Application to the National Health Interview Survey." *Health Services Research* 44: 1622–39.

Executive Office of the President, Office of Management and Budget. 1997. "Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity" [accessed on July 10, 2009]. Available at http://georgewbush-whitehouse.archives.gov/omb/fedreg/1997standards.html

Federal Register. 2010. *Medicare and Medicaid Programs; Electronic Health Record Incentive Program; Proposed Rule.* FR. 75: 1843–2011.

Fiscella, K., and A. Fremont. 2006. "Use of Geocoding and Surname Analysis to Estimate Race and Ethnicity." *Health Services Research* 41: 1482–500.

Geronimus, A. T., J. Bound, and L. J. Neidert. 1996. "On the Validity of Using Census Geocode Characteristics to Proxy Individual Socioeconomic Characteristics." *Journal of the American Statistical Association* 91: 529–37.

Gold, M., A. Dodd, and M. Neuman. 2008. "Availability of Data to Measure Disparities in Leading Health Indicators at the State and Local Levels." *Journal of Public Health Management and Practice* (suppl): S36–44.

Gornick, M. E. 2003. "A Decade of Research on Disparities in Medicare Utilization: Lessons for the Health and Health Care of Vulnerable Men." *American Journal of Public Health* 93: 753–9.

Grieco, E. M., and R. C. Nassidy. 2001. *Overview of Race and Hispanic Origin. Census 2000 Brief.* Washington DC: U.S. Bureau of the Census [accessed on July 10, 2009]. Available at http://www.census.gov/prod/2001pubs/cenbr01-1.pdf

Hasnain-Wynia, R., and D. W. Baker. 2006. "Obtaining Data on Patient Race, Ethnicity, and Primary Language in Health Care Organizations: Current Challenges and Proposed Solutions." *Health Services Research* 41 (4, part I): 1501–18.

Higgins, P. C., and E. F. Taylor. 2009. *Measuring Racial and Ethnic Disparities in Health Care: Efforts to Improve Data Collection.* Policy Brief. Mathematica Policy Research Inc. [accessed on July 10, 2009]. Available at http://www.mathematica-mpr.com/publications/pdfs/health/disparitieshealthcare.pdf

High Value Health Care Project. 2010. "Indirect Estimation of Race and Ethnicity: An Interim Strategy to Measure Population Level Health Care Disparities" [accessed on April 24, 2010]. Available at http://www.healthqualityalliance.org/userfiles/RAND%20issue%20brief%20031810.pdf

Hillenmaier, M., M. Foster, B. Heinrichs, and B. Heier. 2007. "Racial Differences in Parental Reports of Attention-Deficit/Hyperactivity Disorder Behaviors." *Journal of Developmental & Behavioral Pediatrics* 28: 353–61.

Johnson, T. P. 2006. "Methods and Frameworks for Crosscultural Measurement." *Medical Care* 44 (suppl 3): S17–20.

Krieger, N. 1992. "Overcoming the Absence of Socioconomic Data in Medical Records: Validation and Application of a Census-Based Methodology." *American Journal of Public Health* 82: 703–10.

Larson, N. I., M. T. Story, and M. C. Nelson. 2009. "Neighborhood Environments: Disparities in Access to Healthy Foods in the U.S." *American Journal of Preventive Medicine* 36: 74–81.

Leung, B., N. Luo, L. So, and H. Quan. 2007. "Comparing Three Measures of Health Status (Perceived Health with Likert-Type scale, EQ-5D, and Number of Chronic Conditions) in Chinese and White Canadians." *Medical Care* 45: 610–7.

Li, W., J. L. Kelsey, Z. Zhang, S. C. Lemon, S. Mezgebu, C. Boddie-Willis, and G. W. Reed. 2009a. "Small Area Estimation and Prioritizing Communities for Obesity Control in Massachusetts." *American Journal of Public Health* 99: 511–9.

Li, W., T. Land, Z. Zhang, L. Keithly, and J. L. Kelsey. 2009b. "Small Area Estimation and Prioritizing Communities for Tobacco Control Efforts in Massachusetts." *American Journal of Public Health* 99: 470–9.

Madans, J. 2009. *Race/Ethnic Data Collection: Population Surveys and Administrative Records.* Presentation to the Institute of Medicine, Board on Health Services, Subcommittee on Standardized Collection of Race/Ethnicity Data for Healthcare Quality Improvement, February 9, 2009 [accessed on January 22, 2009]. Available at http://www.iom.edu/Activities/SelectPops/RaceEthnicityData/2009-FEB-09.aspx

Marwick, C. 1991. "Hispanic HANES Takes Long Look at Latino Health." *Journal of the American Medical Association* 265: 177–81.

McHorney, C., and J. Fleishman. 2006. "Assessing and Understanding Measurement Equivalence in Health Outcome Measures: Issues for Further Quantitative and Qualitative Inquiry." *Medical Care* 44 (suppl 3): S205–S210.

Meliker, J. R., P. Goovaerts, G. M. Jacquez, G. A. AvRuskin, and G. Copeland. 2009. "Breast and Prostate Cancer Survival in Michigan: Can Geographic Analyses Assist in Understanding Racial Disparities?" *Cancer* 115: 2212–21.

New York City Department of Health and Mental Hygiene. "Designing a New York City Health and Nutrition Examination Survey" [accessed on July 10, 2009]. Available at http://www.nyc.gov/html/doh/html/hanes/section2.shtml

Parlato, J. 2009. *Massachusetts Hospital Discharge Data: Hospital Race and Ethnicity Data Collection.* Presentation at the National Association of Health Data Organizations Webinar, December 17, 2009 [accessed on January 21, 2010]. Available at http://www.nahdo.org/Quality/2010QualityReportingWorkgroup Calls/2009QualityReporting WorkgroupCalls/DecemberCall2009/tabid/156/Default.aspx

P.L. 106–29. Healthcare Research and Quality Act of 1999 [accessed on July 10, 2009]. Available at http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=106_cong_public_laws&docid=f:publ129.106.pdf

P.L. 111–5. American Recovery and Reinvestment Act of 2009 [accessed on July 10, 2009]. Available at http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=111_cong_bills&docid=f:h1enr.txt.pdf

P.L. 111–48. Patient Protection and Affordable Care Act [accessed on July 10, 2009]. Available at http://democrats.senate.gov/reform/patient-protection-affordable-care-act-as-passed.pdf

Prewitt, K. 2005. "Racial Classification in America: Where Do We Go from Here?" *Daedalus* 134: 5–17.

Ramirez, M., M. Ford, A. Stewart, and J. Teresi. 2005. "Measurement Issues in Health Disparities Research." *Health Services Research* 40: 1640–57.

Satcher, D. 1999. "The Initiative to Eliminate Racial and Ethnic Health Disparities Is Moving Forward." *Public Health Reports* 114 (3): 283–7.

Schenker, N., T. E. Ragunathan, P. Chiu, D. M. Makuc, G. Zhang, and A. J. Cohen. 2006. "Multiple Imputation of Missing Income Data in the National Health Interview Survey." *Journal of the American Statistical Association* 101: 924–33.

Schneider, K. L., K. L. Lapane, M. A. Clark, and W. Rakowski. 2009. "Using Small Area Estimation to Describe County-Level Disparities in Mammography." *Preventing Chronic Disease* 6 (4) [accessed on January 18, 2010]. Available at http://www.cdc.gov/pcd/issues/2009/oct/08_0210.htm

Sonnenfeld, N., and J. Sisk. 2008. "Health Care Disparities Research: Using the National Health Care Surveys." *Proceedings of the 2008 Joint Statistical Meetings* 1: 2105–12.

Tang, G. 2009. *Defining Race, Ethnicity, and Language Populations.* Presentation to the Institute of Medicine Committee and Subcommittee on Standardized Collection of Race/Ethnicity Data for Health Quality Improvement, Newport Beach, CA [accessed on July 10, 2009]. Available at http://www.iom.edu/Object.File/Master/64/326/Tang%20Slides.pdf

UCLA Center for Health Policy Research. "California Health Interview Survey. Survey Design and Methods" [accessed on July 10, 2009]. Available at http://www.chis.ucla.edu/design.html

Ulmer, C., B. McFadden, and D. R. Nerenz (Eds.) 2009. *Race, Ethnicity, and Language Data: Standardization for Health Quality Improvement.* Washington, DC: National Academies Press.

U.S. Bureau of the Census. 2008. "A Compass for Understanding and Using American Community Survey Data: What General Data Users Need to Know" [accessed on April 24, 2010]. Available at http://www.census.gov/acs/www/Downloads/ACSGeneralHandbook.pdf

U.S. Department of Health and Human Services. 1985. *Report of the Secretary's Task Force on Black and Minority Health.* Washington, DC: U.S. Government Printing Office.

U.S. Department of Health and Human Services. 2000. *Healthy People 2010. 2nd Edition. With Understanding and Improving Health and Objectives for Improving Health*, 2 vols. Washington, DC: U.S. Government Printing Office [accessed on July 14, 2009]. Available at http://www.healthypeople.gov/Document/tableofcontents.htm

U.S. Department of Health and Human Services. 2003. *National Healthcare Disparities Report.* Rockville, MD: Agency for Healthcare Research and Quality [accessed on July 14, 2009]. Available at http://www.ahrq.gov/qual/nhdr03/nhdr03.htm

U.S. Department of Health and Human Services. 2006. *Healthy People 2010 Midcourse Review.* Washington, DC: U.S. Government Printing Office [accessed on July 10, 2009]. Available at http://www.healthypeople.gov/data/midcourse/default.htm#pubs

U.S. Department of Health and Human Services. 2008. *Phase I Report, Recommendations for the Framework and Format of Healthy People 2020* [accessed on July 10, 2009]. Available at http://healthypeople.gov/HP2020/advisory/PhaseI/default.htm

U.S. Department of Health and Human Services. 2009. *National Healthcare Disparities Report 2008.* Rockville, MD: Agency for Healthcare Research and Quality [accessed on July 14, 2009]. Available at http://www.ahrq.gov/QUAL/nhdr08/nhdr08.pdf

Ver Ploeg, M., and E. Perrin, (Eds). 2004. *Eliminating Health Disparities: Measurement and Data Needs.* Washington, DC: National Academies Press.

Wun, L.-M., T. M. Ezzati-Rice, and R. Baskin. 2008. *Evaluation of the Prediction Model and Oversample of Low Income Persons in the Medical Expenditure Panel Survey.* Proceedings of the Survey Research Methods Section, Joint Statistical Meetings, Denver, 2008 [accessed on January 18, 2010]. Available at http://www.amstat.org/sections/SRMS/Proceedings/y2008/Files/301229.pdf

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.