HIGH-VALUE HEALTH CARE PROJECT
An initiative of the Quality Alliance Steering Committee
Supported by the Robert Wood Johnson Foundation
Directed by the Engelberg Center for Health Care Reform at the Brookings Institution
March 2010

INDIRECT ESTIMATION OF RACE AND ETHNICITY: AN INTERIM STRATEGY TO MEASURE POPULATION LEVEL HEALTH CARE DISPARITIES

Health care equity is a key component of health care quality. In the United States, research has documented variations in care and outcomes based on race, ethnicity, and socioeconomic status. But since race and ethnicity information in health care is currently only available for a very small fraction of the population, far more data must be collected from key health care stakeholders — such as health insurers, and public and private payers — in order to address disparities and prioritize quality improvement efforts.

Indirect estimation methods represent a promising short-term strategy for assessing population-level racial/ethnic health care disparities. While initial indirect estimation algorithms enabled health plans lacking directly collected race and ethnicity data to assess the composition of their membership, they did not work well on all race and ethnicity groups. However, these methods have been significantly refined over the past several years, effectively addressing limitations and improving their accuracy. More recent versions of estimation algorithms, which employ both name analyses and geocoding, appear to substantially increase accuracy. Some of these approaches have been fully validated in several regions of the United States.

This brief addresses some of the important questions and concerns regarding the use of these new methodologies for assessing population-level health care disparities. The discussion herein is based on validation activities of RAND's indirect estimation algorithm.

APPROPRIATE USE OF INDIRECT ESTIMATES

Some health care organizations have used indirect estimation to target individual enrollees or patients. However, these methods should mainly be used to

ACHIEVING HIGHER QUALITY HEALTH CARE

The High-Value Health Care (HVHC) Project is working to make valid, timely, and consistent information about the quality and cost of health care widely available in the United States. The project's diverse collaborative of physicians, nurses, hospitals, health insurers, consumers, employers, government, regional and local initiatives, accrediting agencies, and foundations supports the implementation and use of performance information to:

- Help health care providers improve the quality of patient care;
- Help consumers make informed choices about health care providers;
- Help provide payments that support provider efforts to improve quality and efficiency, rather than simply paying for more intensive treatments; and
- Help reduce racial and ethnic disparities in care.

The HVHC project is supported by the Robert Wood Johnson Foundation and directed by the Engelberg Center for Health Care Reform at Brookings. Learn more at www.healthqualityalliance.org.

broadly ascertain racial and ethnic composition of plan populations in different markets, communities, and large provider groups, as well to estimate potential inequalities in care among these groups.

Indirect estimates can appropriately be used to:

- Estimate racial/ethnic composition;
- Identify disparities in quality of care among different race and ethnicity groups;
- Fill gaps in information where plans have limited self-reported race and ethnicity data;
- Combine with Geographic Information System (GIS) mapping and decision tools to efficiently target community-wide disparities and interventions; and
- Combine with other socio-demographic information (e.g., neighborhood income, linguistic isolation rate) to enhance the indirect estimation results.

A VALID BASIS FOR ACTION

Indirect estimation can be used as an interim strategy to effectively act on population-level disparities immediately. It can also be important for three additional reasons.

First, direct data collection methodologies have often been inadequately implemented. Research documents that even when data on race and ethnicity are collected directly, "eye-balling" patients (at the provider level) often supersedes allowing patients to self-report. Second, research also documents selection bias among plan members or patients in answering demographic questions. For example, people who identify as "White" are more likely to respond to questions about race and ethnicity than are minorities, effectively skewing estimates of racial/ethnic population composition. Third, if a health plan has self-reported race and ethnicity data available on most patients, indirectly estimated data can fill in remaining gaps. In particular, because the advanced indirect estimation process includes geocoding and draws on Census data, a health care organization can simultaneously obtain a variety of other socio-demographic information about their members that they might not otherwise have, such as estimated income, education, and English language proficiency.

Additionally, since the member-level data is "geo-enabled" through the indirect estimation process, a health care organization can take advantage of emerging GIS mapping and decision tools to better target its efforts to reduce disparities.

ADDRESSING CONCERNS ABOUT METHODOLOGY

Reliability of Classification

New methodologies, which use statistical methods to combine an enhanced surname list from the Census Bureau along with geocoding, have proven substantially more reliable than previous methods. In 2008, the U.S. Census Bureau released a new surname list that was far more comprehensive and included unprecedented detail compared to prior lists. This list was based on almost 270 million individuals with valid surnames enumerated on Census 2000, along with self-reported race and ethnicity data.[1] The enhanced list covers 151,671 surnames shared by 100 or more individuals, along with each surname's self-reported racial/ethnic distribution. These names are publicly available and represent 89.8 percent of all individuals enumerated on Census 2000.

Table 1. Percentage of Member Population by Race and Ethnicity Group

Race/Ethnicity	Self-Reported	Indirectly Estimated
Black	6%	5%
Asian[a]	4%	3%
American Indian[b]	0%	1%
Hispanic	17%	16%
White	71%	73%

[a] Asian includes Pacific Islander; [b] American Indian includes Alaskan Native

For each surname with 100 or more occurrences nationally, the publicly available tabulation shows the frequency of occurrence in each of six mutually exclusive categories: (1) Hispanic, (2) White, (3) Black, (4) Asian and Pacific Islander, (5) American Indian/Alaska Native, and (6) Multiracial. Thus for each such name, a highly precise estimate of the probability that people with that name will self-report as belonging to each of the six race and ethnicity groups is provided. It is the use of this new, more detailed surname information, along with the Bayesian method (a means of inferring an unknown), that most distinguishes new indirect estimation approaches from prior ones.[2]

A recent review of the results from applying RAND's indirect estimation algorithm showed a strong correlation between the RAND estimates and self-reported data.

Accuracy of Performance Data

Because the algorithm estimates probabilities that an individual may belong to each of the major race and ethnicity groups, it is not intended to identify an individual's specific race and ethnicity. The probabilities can be used by a health plan to estimate the number or proportion of its members in any given race and ethnicity group (e.g., by adding up those probabilities for each individual and dividing that number by the total population to estimate racial/ethnic composition of an area). The estimates can further be used to assess the rates at which plan members in different race/ethnicity groups are receiving indicated care (e.g., HbA1c test in diabetics).

The review of results, which applied RAND's indirect estimation algorithm, also showed strong correlations between performance rates among self-reported and indirectly estimated measures. Table 2 compares performance rates among members with self-reported data versus indirectly estimated data in estimating HbA1C testing rates.

CONCLUSION

Improved methodologies have made indirect estimation a useful interim strategy to help health plans and other entities determine the racial/ethnic composition of populations. Further, estimated race and ethnicity data can be useful in assessing disparities in care, a key dimension of health care quality. As a substitute for hard-to-get self-reported data, or in combination with such data, indirect estimations can provide health care organizations with valid information on which to base improvement strategies.

Table 2. HbA1C Testing Rate among Diabetics with Self-Reported and Indirectly Estimated Race and Ethnicity

	HbA1C Testing Rate	
Race and Ethnicity Group	Self-Reported	Indirectly Estimated
Black	85%	86%
Asian[a]	93%	94%
American Indian[b]	91%	87%
Hispanic	89%	89%
White	88%	88%

a. Asian includes Pacific Islander; b. American Indian includes Alaskan Native

1. Word DL, Coleman CD, Nunziata R, Kominski R. "Demographic Aspects of Surnames from Census 2000." Available at: www.census.gov/genealogy/www/surnames.pdf (2008) [accessed on July 30, 2008].
2. Elliott MN, Morrison PA, Fremont AM, McCaffrey DM, Pantoja P, Lurie N. "Using the Census Bureau's Surname List to Improve Estimates of Race/Ethnicity and Associated Disparities." Health Services and Outcomes Research Methodology, 2009 9(2): 69-83.