MEDINFO 2007
HISA Ltd Australia
© 2007 The authors. All rights reserved.

Measuring Race and Ethnicity in SNOMED Patient Records

Mary E. Campbell[a]

[a] *Department of Sociology, University of Iowa, United States*

Abstract

One of the challenges facing informatics today is the collection of useful data on racial and ethnic background. The collection of these data is complex and problematic. The current system for coding racial and ethnic data in patient records is often inconsistent and confusing, as an examination of the example of SNOMED illustrates. Principles drawn from the social science research on the topic can be used to improve our methods of collecting racial and ethnic data.

Keywords:

race, ethnicity, Systematized Nomenclature of Medicine (SNOMED), computerized patient record

Introduction: The complexity of "race"

Although "race" and "ethnicity" are social, not biological, constructs, it is nonetheless important that these social groups be measured appropriately in medical informatics. Developing coding systems that use racial and ethnic concepts in a logically consistent way is challenging. Our ideas about measuring race and ethnicity developed in an often haphazard response to nation-specific historical ideas about racial and ethnic difference. We have no internationally comparable or systematic nomenclature for capturing race and ethnicity in health contexts [1-2]. In an effort to contribute to the development of a valid method of coding racial and ethnic groups, I discuss ways that the social scientific evidence on measuring race and ethnicity could be incorporated into informatics systems.

Accurately categorizing "race" and "ethnicity" is not simple. Variation in how these data are collected can significantly change our understanding of the relationship between ethnicity and health. First, some datasets identify patients by appearance, while others allow patients to self-identify. In vital statistics data in the United States, for example, death certificates are often completed based on the appearance of the deceased or other observable factors, which leads to an undercount of Latinos and other minority groups [3].

Second, small differences in the way questions about race and ethnicity are worded or formatted lead to large differences in how individuals identify. For example, many U.S. surveys separate Latino "ethnicity" from the question on "race" [4]. This results in higher rates of identification as "Latino" than questions that include all of the groups in a single question [5]. Survey respondents are also strongly influenced by the examples they are given and the placement of the question [6].

Third, some ethnic groups are simply left out of many data collection efforts. Individuals of Middle Eastern or Arab descent, for example, typically cannot identify that heritage on U.S. health surveys, despite significant health disparities [7].

Results: Racial/ethnic categories in SNOMED

I use SNOMED as an example of a comprehensive clinical terminology to illustrate how race and ethnicity are currently being captured in medical informatics. The January 2006 release of SNOMED [9] codes "ethnic group" and "racial group" data as part of a patient's social context.

The list of "ethnic groups" included in SNOMED is extensive and detailed, ranging from nationalities (e.g. Germans, Dutch), ethnic groups within national boundaries (e.g. Basques, American Indian tribes), religio-ethnic groups (e.g. "Oriental Jews"), panethnic groups that refer to skin color (e.g. "Black") and region of the globe (e.g. "Asian"). Major "racial groups" in SNOMED are based in part on Census terminology [1]. These lists are not based on a single, consistent set of criteria. They include categories based on geography, skin tone, national boundaries, and cultural distinctions.

The separation of "racial groups" from "ethnic groups" is also a potential source of confusion. In the absence of claims based on biological racism, it becomes more difficult to articulate why the categories should be separated in health data [10-11]; for patient records, we simply need to be able to accurately identify the social groups to which the patient belongs.

Discussion: Principles for racial and ethnic data

The social science research on racial and ethnic identification suggests several principles for race and ethnicity vocabulary.

First, *self-identification should be preferred over observer identification*; staff and clinicians should be reminded to *ask* individuals about their ethnicity directly. Self-identifi-

cation is preferable to create consistency across datasets, give clinicians information about the social group that influences the patient's actions, and eliminate the strong influence of observer bias [12-13]. Ethnic self-identification is related to the selection of peer groups, communities, media and cultural outlets. If observer identification is used, it should be identified as such.

Second, *the dichotomy of "race" and "ethnicity" found in the SNOMED classification of ethnicity should be discarded*. Instead, "ethnic origins" should be collected together. As noted above, the inconsistent use of the terms makes their separation nonsensical in most contexts. Racial and ethnic questions that are formatted differently from how patients and staff think have higher rates of inaccuracy and nonresponse [4].

Third, *patients should be allowed to choose more than one ethnic origin*. Multiracial identification is increasing rapidly, increasing the importance of understanding the multiple ethnic origins that influence patients.

Fourth, *patients should also be asked to choose a single ethnic category that they find* ***most*** *representative of their ethnicity.* At times it is necessary to cluster patients into larger groups; it is more meaningful to do this by asking patients about their identification than by arbitrary allocation rules [8, 14-15].

Finally, *patients must be allowed to select from a list of ethnic categories that includes* ***all*** *of the groups that face ethnic discrimination or inequality.* If we hope to target populations that are facing increased health risks due to persecution or differences in socioeconomic opportunities, we cannot exclude meaningful categories. In this respect, the 2006 SNOMED categories are admirable in their level of detail; they include, for example, not simply panethnic categories (e.g. "American Indian") but a detailed list of ethnic categories (e.g. tribes). Conversely, terms that are not socially meaningful for any groups today should be dropped from the list.

It is wise to periodically reassess the ethnic data collected for patients, both because popular understandings of ethnicity change, and because identities are fluid over time [16]. The examples we give patients influence their identification, so it is also important to administer in the questions in a consistent way. National and regional differences will also influence how ethnic labels are understood and interpreted [2].

Conclusion: Recommended question format

I recommend asking patients:

"What are your ethnic origins or ancestry? Please select all the groups that apply." The entries currently found on the detailed list of ethnic and racial groups in SNOMED could be used as the potential answers to the question, once those categories have been pared down to include only categories that are socially meaningful to patients. (For example, the term "Mongoloid" is generally not used as an ethnic group identifier today.) In order to make this rather lengthy list comprehensible to users, it could be organized geographically, by region of ethnic origins.

Individuals who identify more than one ethnic group should also be asked a follow-up question: "Which one of these groups would you say BEST represents your ethnic origins or ancestry?" The same list of potential answers can be used.

The advantages of this format are many. First, questions that combine "ethnic" and "racial" categories into one question have lower rates of nonresponse, improving data quality [4]. The current SNOMED "racial groups" allow Hispanics to belong to a white or black subgroup, but do not do the same for any other "racial group," leading to potential confusion. Second, the level of detail found in the SNOMED ethnic categories is crucial, since panethnic categories such as "Hispanics" and "Asians" mask a tremendous amount of heterogeneity that is important to health outcomes. Third, the follow-up question asking each respondent to choose a single category is essential. Allocating multiracial respondents arbitrarily leads to grouping highly disparate individuals together and incorrectly identifies the category they would have chosen on a single-race survey [8, 14-15]. Using this follow-up question for allocation leads to much better information, since multiracial individuals tend to most resemble the single-race group they choose for themselves.

References

[1] Aspinall PJ. The operationalization of race and ethnicity concepts in medical classification systems: issues of validity and utility. *Health Informatics Journal* 2005; 11 (4): 159-74.

[2] Bhopal, R. Glossary of terms relating to ethnicity and race: for reflection and debate. *J Epidemiol Community Health* 2004; 58: 441-5.

[3] Sandefur GS, Campbell ME and Eggerling-Boeck J. Racial and Ethnic Identification, Official Classifications, and Health Disparities. In: Anderson NB, Bulatao RA and Cohen B, eds. *Critical Perspectives on Racial and Ethnic Differences in Health in Late Life.* Washington, DC: National Academy Press, 2004; pp. 25-52.

[4] Hirschman C, Alba R and Farley R. The Meaning and Measurement of Race in the U.S. Census: Glimpses into the Future. *Demography* 2000; 37: 381-393.

[5] Campbell ME and Rogalin C. Categorical Imperatives: The Interaction of Latino and Racial Identification. *Soc Sci Quart* 2006; 87 (5): 106-28.

[6] Rosenwaike I. Ancestry in the United States Census, 1980-1990. *Soc Sci Res* 1993; 22 (4): 383-90.

[7] Lauderdale DS. Birth outcomes for Arabic-named women in California before and after 9/11. *Demography* 2006; 43 (1): 185-201.
[8] Mays VM, Ponce NA, Washington DL and Cochran SD. Classification of Race and Ethnicity: Implications for Public Health. *Annu Rev Publ Health* 2003; 24: 83-110.
[9] SNOMED. http://www.snomed.org/.
[10] Hirschman C. The Origins and Demise of the Concept of Race. *Popul Dev Rev* 2004; 30 (3): 385-415.
[11] Cornell S and Hartmann D. *Ethnicity and Race: Making Identities in a Changing World.* Thousand Oaks, CA: Pine Forge Press, 1998.
[12] Hill ME. Race of the Interviewer and Perception of Skin Color: Evidence from the Multi-City Study of Urban Inequality. *Am Sociol Rev* 2002; 67: 99-108.
[13] Jones EE, Wood GC and Quattrone GA. Perceived Variability of Personal Characteristics in In-Groups and Out-Groups: The Role of Knowledge and Evaluation. *Pers Soc Psychol B* 1981; 7: 523-8.
[14] Parker JD, Schenker N, Ingram DD, Weed JA, Heck KE and Madans JH. Bridging between two standards for collecting information on race and ethnicity: an application to Census 2000 and vital rates. *Public Health Report* 2004; 119: 192-205.
[15] Schenker N and Parker JD. From single-race reporting to multiple-race reporting: using imputation methods to bridge the transition. *Statistics in Medicine* 2003; 22: 1571-87.
[16] Harris DR and Sim JJ. Who is Multiracial? Assessing the Complexity of Lived Race. *Am Sociol Rev* 2002; 67 (4): 614-27.