Measuring Race and Ethnicity in SNOMED Patient Records

Mary E. Campbell
Department of Sociology, University of Iowa

Measuring Race and Ethnicity

- *How might the social science evidence on the measurement of race and ethnicity be useful for the design of informatics systems?*
  - This study uses SNOMED as an example of a comprehensive clinical terminology to illustrate how race and ethnicity are currently captured in medical informatics, and discusses how recent research on the measurement of race and ethnicity can be distilled into five useful principles that can guide measurement of race and ethnicity, leading to better data quality.

The complexity of classifying "race" and "ethnicity"

- Measurement of race and ethnicity is very sensitive to question format
  - We have no internationally comparable or systematic nomenclature for capturing race and ethnicity in health contexts [1-2].
  - Self-reported race & ethnicity is different from the race & ethnicity that observers assign to patients [3-5].
  - Minor differences in the wording of questions leads to major differences in self-identification [6-8].

Racial and Ethnic Categorization in SNOMED

- The January 2006 release of SNOMED codes "ethnic group" and "racial group" data as part of a patient's social context.
  - The list of "ethnic groups" is extensive, 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 [6, 8]. They are: African, American Indian/Alaska Native, Asian or Pacific Islander, Australian Aborigine, Caucasian, Hispanic, Indian, and Mixed Racial Group (with subgroups included under several of these groups).

Racial and Ethnic Categorization in SNOMED, continued

- Measurement of race and ethnicity has important implications for health data [9-10].
- These lists of groups are not created based on a single, consistent set of criteria. They include categories based on geography, skin tone, national boundaries, and cultural distinctions [11-12].
- These contradictions are potentially confusing. For example, the “racial groups” allow Hispanics to belong to a white or black subgroup, but do not do the same for any other “racial group” [6].

Principles for racial and ethnic data

1. *Self-identification should be preferred over observer identification*; staff and clinicians should *ask* individuals about their ethnicity. This is preferable for several reasons:
   a. Creates consistency across data sources,
   b. Gives clinicians information about the social group that actually influences the person's identity and actions, and
   c. Eliminates the strong influence of observer bias [3-5]. (An observer's own racial identification strongly influences estimates of another person's ethnic group.)

Principles for racial and ethnic data, continued

2. *The dichotomy of "race" and "ethnicity" found in the SNOMED classification of ethnicity should be discarded*.

"Ethnic origins" should be collected together. The separation of "racial groups" from "ethnic groups" is often a source of confusion. In the absence of claims based on biological racism, it becomes difficult to articulate why the categories should be separated in health data [11-12]; for patient records, we simply need to be able to accurately identify the social groups to which the patient belongs. Racial and ethnic questions that are formatted differently from how patients and staff think of themselves have higher rates of inaccuracy and nonresponse [6].

Principles for racial and ethnic data, continued

3. *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.

4. *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 [10, 13–15].

Principles for racial and ethnic data, continued

5. *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 [9]. In this respect, the 2006 SNOMED categories are admirable in their level of detail.

Recommended question format

- “What are your ethnic origins or ancestry? Please select all the groups that apply.”

- 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?”

Advantages of this format

1. Questions that combine “ethnic” and “racial” categories into one question have lower rates of nonresponse, improving data quality [6].
2. 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.
3. The follow-up question asking each respondent to choose a single category is essential, because multiracial individuals tend to most resemble the single-race group they choose for themselves. 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 [10, 13-15].

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. 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.
5. 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.
6. 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.
7. Campbell ME and Rogalin C. Categorical Imperatives: The Interaction of Latino and Racial Identification. *Soc Sci Quart* 2006; 87 (5): 106-28.
8. Rosenwaike I. Ancestry in the United States Census, 1980-1990. *Soc Sci Res* 1993; 22 (4): 383-90.
9. Lauderdale DS. Birth outcomes for Arabic-named women in California before and after 9/11. *Demography* 2006; 43 (1): 185-201.
10. 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.
11. Hirschman C. The Origins and Demise of the Concept of Race. *Popul Dev Rev* 2004; 30 (3): 385-415.
12. Cornell S and Hartmann D. *Ethnicity and Race: Making Identities in a Changing World.* Thousand Oaks, CA: Pine Forge Press, 1998.
13. 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.
14. 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.
15. Harris DR and Sim JJ. Who is Multiracial? Assessing the Complexity of Lived Race. *Am Sociol Rev* 2002; 67 (4): 614-27.

Acknowledgements: I would like to thank Dr. James Campbell at the University of Nebraska Medical Center and three anonymous reviewers.

Contact information: Mary E. Campbell, Department of Sociology, University of Iowa, W140 Seashore Hall, Iowa City, IA 52242, United States. Email: mary-e-campbell@uiowa.edu. Fax: (319) 335-2509, Phone: (319) 335-2495.