

Appendix I

Subcommittee Template: Developing a National Standard Set of Spoken Language Categories and Coding

The Subcommittee on Standardized Collection of Race/Ethnicity Data for Healthcare Quality Improvement recommends the collection of the variable spoken “language need” for each individual. Language need is to be assessed through two questions: the first is an individual’s personal assessment of his/her ability to speak English and then their preferred spoken language for a health-related encounter (Recommendation 4-1). 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 defines limited English proficiency for health care purposes as someone who speaks English less than very well.

When data are shared from one entity to another (e.g., providers to health plan or health plans to states), standardization helps ensure that data can be combined for like categories. Unlike race and Hispanic ethnicity, there is no Office of Management and Budget (OMB) standard for language categories, thus the subcommittee recommends that the Department of Health and Human Services (HHS) develop national standard lists of spoken and written languages and codes (Recommendation 6-1a) and that entities choose their categories from the national standard list (Recommendation 4-3) according to the needs of the population they serve or study (Recommendation 4-2). When a health care entity designs its collection instruments, whether paper or electronic, it may, because of space considerations, have to use a limited number of pre-selected response categories. Therefore, such a response list should always include an “Other, please specify: ___” option to ensure collection of each person’s language need (Recommendation 4-2). Some electronic data collection systems are more sophisticated, and by using keystroke recognition can accommodate hundreds of languages.

DEVELOPMENT OF THE TEMPLATE

The subcommittee did not identify a single existing category list that it believed was ready to serve as a national standard set. To develop a template of spoken languages spoken in the United States, the subcommittee compiled the attached list to serve as a draft template of language names and coding possibilities. (An online searchable and sortable Excel version of the list is available at: www.iom.edu/datastandardization.) Census Bureau data on languages spoken at home was a logical place to start to compile lists of languages; the Census has compiled

approximately 530 language names corresponding to about 380 language codes.¹ Some of these languages are nearing extinction. Another group, International Organization for Standardization (ISO) has established code sets for thousands of languages; the ISO language lists and particularly their coding focus on distinct languages with distinct codes, whereas the Census Bureau is more likely to give related languages the same code. The ISO codes have evolved from a first-generation two-letter coding system (ISO 639-1), to a three-letter system to accommodate additional languages primarily for bibliographic uses (ISO 639-2), to a set that now incorporates more three-letter codes to cover 6,000 known languages in the world (ISO 639-3). The ISO 639-3 codes are intended “to provide a comprehensive set of identifiers for all languages for use in a wide range of applications, including linguistics, lexicography and internationalization of information systems.”²

The subcommittee list began with the Census Bureau’s summary file 3 (SF3) technical documentation list of approximately 530 languages and 380 three digit numerical codes;³ these are presented in the first two columns. Names that are not in all caps are considered to have a relationship to an ALL CAPS language name and receive the same code.⁴ The Census Bureau could not confirm whether persons speaking the ALL CAPS languages would be understood by those with the same code; the online Excel file can be sorted by the code number to see which languages have overlapping codes. Additional language names, not on the Census list, were added to the Census names column based on previous surveys conducted by Hospital Research & Educational Trust (HRET) of a representative sample of hospitals and the National Association of Community Health Centers of a representative sample of health centers;⁵ requests to Language Line, an interpretation and translation service;⁶ and subcommittee collection of additional names from a handful of providers.⁷ The languages added to the initial Census list are indicated by an * next to the Census code number; the code number assigned was provided by Census Bureau staff to indicate how they would have coded the response; some remain uncoded.⁸ This resulted in approximately 650 total language names, of which approximately 300 were identified as being used in a health care context. A column was added to indicate categories for which the Modern Language Association reports there were responses in Census 2000;⁹ the subcommittee ran Census PUMS data but did not find any further languages since languages with smaller numbers of persons reporting the language were aggregated together.

Each language in the first column was then matched to different generations of ISO codes which are alphabetic rather than numeric codes. ISO 639-2 codes are maintained by the Library of Congress and are coded as two letters; the ISO 639-3 codes are three letter codes currently maintained by SIL International. ISO codes start with the most comprehensive set (ISO-639-3); after the codes, the language name under the ISO categorization scheme is listed.

The names of languages often have multiple possible spellings, even between the Census Bureau and ISO 639 language lists there are alternate spellings, and patients may provide an alternative spelling as well. The column,

¹ U.S. Census Bureau. 2007. Census 2000 Summary File 3—Technical documentation. Appendix G language code list. Washington, DC: U.S. Census Bureau.

² 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).

³ The Census Bureau included the notation n.e.c. next to a language name to means not elsewhere categorized. Some of the languages that may have fallen into these categories may now be listed in column A due to the additions the subcommittee made to the list of languages.

⁴ U.S. Census Bureau. 2002. *Census 2000 summary file 3: Technical documentation*. <http://www.census.gov/prod/cen2000/doc/sf3.pdf> (accessed August 3, 2009).

⁵ Hasnain-Wynia, R., J. Yonek, D. Pierce, R. Kang, and C. H. Greising. 2006. *Hospital language services for patients with limited English proficiency: Results from a national survey*. Chicago, IL: Health Research and Educational Trust (HRET)/AHA; National Association of Community Health Centers. 2008. *Serving patients with limited English proficiency: Results of a community health center survey*. Bethesda, MD: National Association of Community Health Centers and National Health Law Program.

⁶ Language Line Service. 2009. *List of languages by Language Line Services*. <http://www.language-line.com/page/languages/> (accessed June 12, 2009).

⁷ Personal communications from Emilio Carrillo, New York Presbyterian Hospital, May 11, 2009; Alice Chen, San Francisco General Hospital, July 7, 2009; Maria Moreno and Traci Van, Sutter Health, July 22, 2009; Shiva Bidar-Sielaff, University of Wisconsin Health, May 11, 2009.

⁸ Personal communication, H. Shin, U.S. Census Bureau, July 13, 2009.

⁹ Modern Language Association. 2009. All languages reported to the U.S. Census in 2000. http://www.mla.org/map_data_langlist&mode=lang_tops (accessed May 26, 2009).

Other Names and Additional Information includes some of the alternate spellings and names the subcommittee encountered, but these should not be considered all of the alternative names. Languages might even be called slightly different names, such as Amish, Pennsylvania Dutch, or Pennsylvania German, or be quite different. This need not be a barrier to the list of choices developed locally as long as it is clear on a national standard list how to categorize the alternative spellings or names.

Since the Census language list included names that appeared in responses to earlier censuses, some languages were thought to be no longer in use. The American Indian and Alaska Native languages were reviewed to determine whether they were now considered extinct or nearly extinct by Ethnologue which tracks the world's living languages. Ethnologue uses the term nearly extinct when "only a few elderly speakers are still living."¹⁰ Notations are made of this status in the column titled, Other Names and Additional Information; approximately 80 of the 650 languages were identified as extinct or nearly extinct.

ADAPTATION OF THE TEMPLATE TO A NATIONAL STANDARD LIST

Arriving at the possible names for a national category list appears fairly straight forward; the accompanying list is likely to identify most that will be encountered. Changes in immigration patterns over time may result in additional names; thus, category and code lists will have to be maintained (Recommendation 6-1a). Deciding on which coding scheme to adopt is more challenging. In its incidental collection of information on languages, the subcommittee encountered more instances of use of the ISO coding scheme; however, the Census has data on languages spoken at home and the degree of limited English proficiency for many languages that entities use to learn about the populations in their service areas. The subcommittee believes that there are advantages to both the Census Bureau and ISO coding schemes for languages, and does not endorse one over the other. The subcommittee indicates the need for HHS to consult with the Census Bureau, the registration authorities for the ISO codes, and others that establish unique coding for interoperability, such as HL7.

If the Census coding approach were to be adopted, the subcommittee notes that the Census list of languages and codes would likely need some additional changes to be useful. The primary limitation of the Census Bureau coding scheme is that it uses the same code for multiple related languages, while the ISO list has unique codes for each language. To the extent that patients who are not English proficient need language assistance services in distinct languages in order to facilitate understanding during patient-provider interactions, a care provider's ability to track specific languages would be enhanced by unique coding for distinct languages; this could happen by either expanding the Census codes or adopting the detail of the ISO codes. Currently, there is no specific Census code for English. Sign language, an important communication tool, is not a unique language response on the Census, and generally would code the person as speaking English. By contrast, ISO-639 has unique codes for 130 types of sign languages. For health care purposes, some entities have found a separate category noting which persons have speech loss has been useful to understand the communication needs of all patients. Further options for "declined," "unavailable," or "unknown" are also useful when data are being recorded to determine the portion of the service population from whom language data have been collected.

The subcommittee did not generate a list of written languages, but illustrates these needs with the experiences of Kaiser Permanente and Contra Costa Health Plans in appendixes G and H. (Contra Costa used ISO two-letter codes supplemented by their own local coding). The ISO codes represent both spoken and written language names; separate script codes apply to written languages, as well, to describe their lettering.¹¹

¹⁰ Ethnologue. Endangered languages. http://www.ethnologue.com/nearly_extinct.asp (accessed July 20, 2009).

¹¹ SIL International. 2009. *Scope of denotation for language identifiers*. <http://www.sil.org/iso639-3/scope.asp> (accessed July 20, 2009).