Processing of Race and Ethnicity in the National Survey on Drug Use and Health

Eric Grau, Peilan Martin, Peter Frechtel, Jeanne Snodgrass, and Rachel Caspar

Joint Statistical Meetings

Minneapolis, MN August 7-11, 2005
NSDUH contains data about drug use and health from 70,000 respondents across all 50 states and the District of Columbia.

The NSDUH sample design is multistage cluster design with a predefined number of strata within each state and D.C.

Race and ethnicity information is collected for the purposes of defining domains summarizing drug use patterns in analysis tables produced from NSDUH data.
1977 Statistical Policy Directive #15 on race/ethnicity
- Two ethnicities: Hispanic and non-Hispanic
- Four race levels: white, black, American Indian/Alaska native, Asian/Pacific Islander
- Each individual defined by one race category

1997 revisions to OMB standards
- Two ethnicities: Hispanic and non-Hispanic
- Five race levels: white, black/African American, American Indian/Alaska native, Asian, Native Hawaiian & Other Pacific Islander
- Each individual defined by one or more race categories
Prior to 1999, NSDUH race questions complied with the 1977 OMB directive.

Major redesign in 1999 provided opportunity to change race questions.

From 1999 onwards, NSDUH race/ethnicity questions complied with 1997 OMB revised standards.
Which of these groups describes you? Just give me the number or numbers from the card.

TO SELECT MORE THAN ONE CATEGORY, PRESS THE SPACE BAR BETWEEN EACH CATEGORY YOU SELECT.

- 1 WHITE
- 2 BLACK / AFRICAN AMERICAN
- 3 AMERICAN INDIAN OR ALASKA NATIVE (AMERICAN INDIAN INCLUDES NORTH AMERICAN, CENTRAL AMERICAN, AND SOUTH AMERICAN INDIANS)
- 4 NATIVE HAWAIIAN
- 5 OTHER PACIFIC ISLANDER
- 6 ASIAN (FOR EXAMPLE: ASIAN INDIAN, CHINESE, FILIPINO, JAPANESE, KOREAN, AND VIETNAMESE)
- 7 OTHER (SPECIFY)
[IF ASIAN] Which of these Asian groups describes you? Just give me the number or numbers from the card.

TO SELECT MORE THAN ONE CATEGORY, PRESS THE SPACE BAR BETWEEN EACH CATEGORY YOU SELECT.

- 1 ASIAN INDIAN
- 2 CHINESE
- 3 FILIPINO
- 4 JAPANESE
- 5 KOREAN
- 6 VIETNAMESE
- 7 OTHER (SPECIFY)
Categories Used for Editing and Imputation

- White
- Black/African American
- American Indian/Alaska Native
- Native Hawaiian
- Other Pacific Islander
- Asian Indian
- Chinese
- Filipino
- Japanese
- Korean
- Vietnamese
- Other Asian
Tracked Race Information

- Single race responses
- General categories used for imputation (e.g., nonwhite)
- Multiple race responses
  - A single multiple race level combining the 5 OMB categories
  - Specific 2-, 3-, and 4-way combinations among the 4 1977 OMB categories
  - A single level for multiple Asian responses
  - A single level for Native Hawaiian & Pacific Islander
- Indicators for white, black, American Indian/Alaska native, Asian, Native Hawaiian, and Other Pacific Islander
Editing of Race

- Respondent confusion about what is considered a race
  - Considerable missing data
  - Most missing data concentrated among Hispanics
  - NSDUH tables don’t use race information about Hispanics
  - Race used in editing of Hispanicity and vice versa

- Editing involved
  - Coding write-in responses (3 potential other-specify responses)
  - Setting up imputation classes for cases where specific race not clear
Mapping of Race Other-Specifies

- Interviewers write in responses (called “other-specifies”) if respondents feel given categories do not apply to them.

- Roughly 1% of total responses were definitively determined using write-in information, with an additional small percentage providing information that would limit possible imputed values.

- Write-in responses are assigned codes in a data “dictionary” that is updated each quarter.
  - Responses such as “Puerto Rican” and “Porto Rico” would be given the same code.
  - Responses such as “Arab” would be given a special code, which would later be defined as “white.”
Four general types of other-specify codes

- Directly mapped codes
  - Racial category codes (e.g. “mestizo”)
  - Geographic category codes (e.g., “Polish”)

- Indirectly mapped codes (e.g. “Bolivian”)
  - Used for most racially heterogeneous countries
  - Race randomly determined using census data of country of origin

- Codes informative for imputation (e.g. “brown”)
- Noninformative codes

- Most Hispanic responses treated separately
Mapping of Race Other-Specifies

- If two or more countries are given:
  - Do random assignments for each country based on census data, then combine the information

- If a race (from a listed category) and a country (from the other-specify) are given
  - Ignore the country and use the race

- Other situations are handled on a case-by-case basis.
Assigning Single Race for Multiple Race Respondents

- Imputation models for race did not include a category for multiple race
  - Multiple race category small and very heterogeneous
  - Number of categories in response variable kept to a minimum: white, black, American Indian/AK native, Asian/Pacific Islander
- Others also had need for variable comparable to that created prior to 1999
- From 1999-2002, a question was available where multiple race respondents could select the race they “most” identified with.
- What to do in 2003 and 2004?
Assigning Single Race for Multiple Race Respondents

- Use information from 2000-2002 to predict single race among multiple race respondents in 2003-2004

- Eleven logistic models were fit using data from 2000-2002
  - Each model included a type of multiple race respondent (e.g., black and white)
  - The response variable in the model: race that multiple race respondents indicated they “most” identified with
  - Parameter estimates from these models were applied to data from 2003 and 2004

- Single race for 2003 and 2004 determined by random assignment using predicted values from models
Imputation of Race

- Imputation conducted using Predictive Mean Neighborhood (PMN) imputation method
- PMN also applied to other variables in NSDUH. It is a combination of predictive mean matching and nearest neighbor hot deck
- PMN is described in detail in Singh, Grau, and Folsom (2001)
PMN attributes:

- Model fitted among respondents with complete data and applied to item nonrespondents (missing race)
- Mahalanobis distances calculated between vectors of predicted means for all recipients (item nonrespondents) and donors (item respondents)
- Neighborhoods of potential donors determined by taking closest Mahalanobis distances for each recipient
- For each recipient, donor randomly selected from neighborhood of potential donors
Imputation of Race

- Imputation Model was a multinomial logistic model with 4 levels: White, Black, American Indian/AK native, Asian/Pacific Islander
- Race of multiple race respondents determined using single race model given earlier
- Donors and recipients matched on predicted means from above model; if donor happened to be multiple race, then recipient was also multiple race
Imputation of Race

- Special status for Mexicans, Puerto Ricans, Cubans, Central/South Americans who did not list a country, and Hispanics with no Hispanic group
  - Instead of using census data, race for respondents from these countries were imputed.
  - Donors were limited to be from the same Hispanic region

- Partial information from recipient was used to limit pool of donors in some cases (e.g., brown, Asian nonspecific, etc.)
  - Probabilities in predictive mean vector were made conditional on what was known
Race Breakdown, 2004 (percent)
Source of Value If Not from Given Levels in Questionnaire, 2004 (percent)
Concluding Remarks

- Race is a nebulous quantity, and self-identification can change over time depending upon the context.
- Adding the possibility of choosing more than one race greatly increases the complexity, increasing the possibility of a “moving target” when identifying race.
- Nevertheless, race is an important variable when defining domains for measuring drug use and health.
Acknowledgments

This research was funded by a contract from Substance Abuse and Mental Health Services Administration (SAMHSA), Office of Applied Studies.
References


Contact Information

- Eric Grau
  - Research Triangle Institute
  - E-mail: egrau@rti.org  Phone: 215-321-3947

- RTI website for JSM
  - http://www.rti.org/jsm