

2006-2008 General Social Survey Panel Validation

Tom W. Smith

February, 2009

GSS Methodology Report No. 111

In 2008 attempts were made to reinterview 2,000 respondents from the 2006 General Social Survey (GSS). Of these 40 were determined to have died and at least 19 either no longer resided in the United States or no longer lived in a household. Of the eligible 1,941 respondents from 2006, 1,536 were interviewed and 405 were nonrespondents.

In attempting to contact and interview the 2006 respondents, interviewers had respondent's name, address, and (usually) phone number from the 2006 questionnaire and from interim mailings and databases searches used to track respondents who had moved. For a minority of cases, information on social security number and email address were also available. They also had the name and contact information for a person who was not the member of the respondent's household, but who would know where they lived. In addition to the contact information, preloads included data on the respondent's age, gender, race, ethnicity, and language interviewed in in 2006.

To locate the respondent interviewers were instructed to personally visit the best known address for the respondent (based on the 2006 GSS or interim updates). If the respondent no longer resided at that address, interviewers are instructed to find their current address by a) talking to their former neighbors, b) contacting the person listed as knowledgeable about respondent's residence, c) asking the field manager to use a known email address and utilize search databases to locate the respondent, and d) if the field manager was unsuccessful, then search specialists in the central office locating shop conducted database searches.

Since interviewers have the name, contact information, and demographics of the 2006 GSS respondents, there is generally little problem in identifying the correct respondent. But there is always a possibility of having someone other than a 2006 GSS respondent wrongly interviewed. Such potential errors could occur because of family members having the same name (e.g. Juniors and Seniors), some people were located through databases without a direct trail to the last known verified address/phone number, the respondent's limited language or cognitive ability, or other reasons.

To check whether the 1,536 reinterviews were actually of all with 2006 GSS respondents, an analysis of 2006 and 2008 demographics was carried out. For example, gender in 2006 and 2008 were compared. This showed 14 discrepant cases. These cases were examined in detail and based on the respondent's first name and gender-identifying words in the household enumeration forms (HEFs) in 2006 and 2008, the correct gender of these 14 cases was determined. This indicated eight miscodes in 2006 and six in 2008. All appear to be simple, random mispunches and the level of miscodes is consistent with that previously documented (Tom W. Smith, "Discrepancies in Gender Codes," GSS Methodological Report No. 102. Chicago: NORC, 2005). In a subsequent analysis, panel reliability and stability in general for 2006-2008 will be examined (as in Tom W. Smith

and C. Bruce Stephenson, "An Analysis of Test/Retest Experiments on the 1972, 1973, 1974, and 1978 General Social Surveys," GSS Methodological Report No. 8. Chicago: NORC, 1979).

Over all agreements across the panel waves were examined for the following variables: a) gender, b) year of birth, c) country of birth, d) state of residence at age 16, e) race, and f) family of residence at age 16. This comparison identified four cases with multiple disagreements that could not be plausibly explained by a simple coding error (e.g. transposing year of birth) or a reasonable alternate way of reporting a fact (e.g. being listed as white and black at one time and black and white at the other time).

The 2006 and 2008 HEFs for these cases were examined and additional variables such as education, marital status, and household composition were checked. Most discrepancies were clearly due to an error in entering a fact such as year of birth or gender at one point in time (e.g. as was the case for all 14 gender differences). For some discrepancies the correct response was uncertain and did not appear to represent a simple miscode (e.g. reporting state of residence at age 16 as code 23 (Pennsylvania) vs. code 64 (Mississippi)). But in no case was there any strong evidence of a wrong person being reinterviewed.

Additionally, there could be wrong "respondents" due to falsified interviews. As with the cross-sectional cases, these of course are guarded against by the usual NORC validation procedures of monitoring interviewers and recontacting respondents. In addition, in 2008 computer assisted recorded interviews were also conducted and used to valid interviews. No validation problems were detected.

In sum, while numerous discrepancies occur between 2006 and 2008 reports of the same unchanging facts, these differences appear to represent measurement error and are not the result of wrong respondents being reinterviewed.