Six Alternative Weights that Adjust for Attrition in the 2006-2014 General Social Survey Panels[*]

Stephen L. Morgan
Johns Hopkins University

Jiwon Lee
Johns Hopkins University

GSS Methodological Report No. 132

THE GENERAL SOCIAL SURVEY
GSS

June 2020

INTRODUCTION

From 2006 to 2014, the GSS was designed to be a biennial three-wave panel. The three cross-sectional samples drawn in 2006, 2008, and 2010 were reinterviewed two and four years later.[1] This design yielded a "treble panel" that includes three panel datasets of three waves each, with overlapping panels from 2008 through 2012 (see Tables 1 and 2 below). While panel datasets are common, overlapping panel datasets of separate samples of the full adult population are rare and present unique analysis possibilities.

In this report, we present strategies for constructing weights to adjust for attrition in the GSS treble panel. We offer Stata code for the construction of the weights that we explain, as well as data files of weights that researchers may wish to adopt for their own use. These files can be downloaded from GitHub (URL here) and the Open Science Framework (URL here).

Our goal is not to provide "official" GSS weights, or even weights that we regard as definitive. Weight construction for panel surveys with attrition is best understood as a modeling challenge that researchers should confront directly. Readers are welcome to extend, tailor, or eschew our weighting approach, given the needs of their own projects.

In the remainder of this report, we first explain the panel design of the GSS as well as the contours of the realized samples. We then present the models that we used to estimate predicted probabilities of attrition for all baseline sample members. We offer six sets of estimated probabilities, which vary by estimation sample and the target population of interest. We then explain the GSS weights developed by NORC, and we show how our estimated probabilities of attrition can be used to extend the NORC weights so that they adjust for attrition. In conclusion, we discuss alternative approaches to the one we have chosen.

PANEL STRUCTURE AND TARGET POPULATION

Since 2006, the target population of each GSS cross-sectional sample has been the adult English and Spanish speaking population living at residential addresses in the United States.[2] Between 2006 and 2014, the treble panel extended this design to enable the investigation of within-individual change as well as the reliabilities of GSS measures (see Smith and Schapiro 2017; Hout and Hastings 2016).

---

[1] The design was discontinued because of a mid-grant cut in NSF funds, passed through as a response to the Federal budget crisis that resulted in a FY 2013 "sequester" (the result of which was a 5 percent cut in FY 2013 appropriations to NSF). When asked by NSF to provide a recommendation on what to cut for the 2014 fielding of the GSS, a decision was taken, in coordination with NSF and the GSS Board of Overseers, to retain the sample size for the 2014 cross-sectional GSS and instead cancel the 2014 re-interviews for the 2012 cross-section. Enough funds remained to complete the 2014 re-interviews of the 2010 cross-sectional sample.

[2] Prior to 2006, and thus before the treble panel was conducted, the survey instrument was not offered in Spanish. During these years, the implicit target populations were composed only of English speakers.

Realized Sample Sizes and Cumulative Rates of Reinterview

Table 1 presents the realized sample sizes by year for each wave of each panel.[3] The final column presents the cumulative percent reinterviewed for each panel. Those not reinterviewed are those who attrited from the study, and Table 1 indicates that slightly more than one-third of each base-year panel sample attrited within four years.

Table 1. Realized panel sample sizes and cumulative percentages reinterviewed

Panel base year	Year of Interview					Weighted cumulative percent reinterviewed
	2006	2008	2010	2012	2014	
2006	2,000	1,536	1,276			65.4
2008		2,023	1,586	1,295		63.4
2010			2,044	1,551	1,304	63.9

Notes: The cumulative percentage reinterviewed is weighted by the base-year NORC weight *wtssnr*.

Table 1 also makes clear that base-year respondents had two opportunities to attrite: at the point of the first reinterview and then at the point of the second reinterview. In addition, respondents who attrited at the point of the first reinterview were dropped from the GSS panel by NORC and thus deemed attriters for the second reinterview as well.

For our work in this report, we will develop weights that are appropriate for projects that utilize data collected for those who never attrited. In particular, we will develop weights for use when modeling the outcomes for respondents who participated in all reinterview waves (1,276 for the first panel, 1,295 for the second panel, and 1,304 for the third panel). For projects that use only the data from the base-year wave and the first reinterview wave, slight modifications to our strategy could be adopted.

The Complications Produced by Ineligibility for Reinterview

When a respondent attrites from the reinterview sample, it is often possible for the field staff to determine why. For example, the contact identified by the respondent in the base-year wave can inform the field staff in a subsequent year through the recontact process that the respondent became ineligible for the study, most commonly because of death or emigration from the United States.[4] These types of attriters are then classified as "out of scope" for the panel because they

---

[3] Respondents in the 2006, 2008, 2010 cross-sectional samples were designated as respondents in the base-year samples of the treble panel. For 2006, a random subset of 2000 respondents from the full cross-sectional sample was selected for panel follow-up. For 2008 and 2010, all respondents from the cross-sectional samples were selected for follow-up. These respondents were approached for reinterviews two and four years later, concurrent with the interviews of the new GSS cross-sectional samples in 2008 through 2014.

[4] A third reason for ineligibility, moving from a residential address to institutionalized quarters, or simply no longer residing in a residential housing unit, is less common. For simplicity, we will not discuss these types of cases within the text of this report.

are not members of the target population of English and Spanish speakers living at residential addresses in the United States for all three waves of the panel.[5]

Table 2 has the same structure as Table 1, but it excludes all individuals who transitioned to ineligibility. For example, the 2006 base-year sample included 2000 respondents, but 109 respondents were subsequently classified as ineligible for one or both follow-up waves. These individuals are dropped from the first row of Table 2. When restricted in this fashion, the rate of attrition declines slightly for all three panels.

Table 2. Panel sample sizes and cumulative percentage reinterviewed, among respondents who were not determined to be ineligible for either the second or third wave

| Panel base year | Year of Interview | | | | | Weighted cumulative percent reinterviewed |
	N in 2006	N in 2008	N in 2010	N in 2012	N in 2014	
2006	1,891	1,487	1,276			68.6
2008		1,918	1,539	1,295		66.1
2010			1,923	1,504	1,304	67.5

Notes: See Table 1.

A comparison of Tables 1 and 2 reveals a crucial set of complications that can only be resolved based on an investigator's choices in a given project. Should an attrition-adjustment weight attempt to align the realized panel sample to the realized base-year sample that generalizes to the base-year target population? Or, should the attrition-adjustment weight attempt to align the realized panel sample to the subset of the realized base-year sample that remained eligible for all three waves of the panel (and thus generalizes only to a stable core of the base-year target population that could be sampled for the GSS two and four years later)? In other words, should the attrition-adjustment weight attempt to align the realized panel sample to respondents in the leftmost diagonal of Table 1 or Table 2?

In this report, we will offer two sets of weights. For some projects on outcomes related to ineligibility transitions (e.g., support for health spending or attitudes toward levels of immigration), results could differ based on which set of weights is utilized, even though the exact same respondents would be included in the panel-based models themselves (i.e., the rightmost diagonals of Tables 1 and 2 are the same). The difference would arise from how much additional weight might be given to panel members with characteristics similar to base-year respondents who died or left the country before the reinterviews were conducted. We will show below that these two sets of weights differ slightly.

---

[5] It should be noted that the status "out of scope" may be slightly undercounted among those who attrite. Although extensive effort was devoted to tracing respondents in order to determine their status, it is likely that some individuals who attrited in each panel became genuinely out of scope and that NORC was unable to definitively determine that this was the case.

For example, the individual with the largest value for the weight that generalizes to the realized base-year sample is a non-citizen, Spanish speaker, living in the West South Central census region, who completed her two reinterviews in her nineties.[6] Her estimated weight is larger for one set of weights because that set of weights uses her to represent all observationally similar base-year respondents, including those who became ineligible. Implicitly, she stands in simultaneously for older respondents who were more likely to die and other respondents who had a higher chance of emigration. Her weight for the alternative set of weights – the set that generalizes only to the always-in-scope subset of the realized base-year sample – is smaller because that set of weights uses her to represent only base-year respondents who remained eligible through the final panel reinterview.

ESTIMATED PROBABILITIES OF ATTRITION

For the approach taken in this report, the first step in constructing a weight that adjusts for non-random attrition is to estimate each base-year respondent's probability of attrition. Smith and Son (2010) and Frankel and Hillygus (2014) have provided an initial analysis of predictors of attrition for the 2006 GSS panel, and we build on their findings for our specifications below. Their insight, affirmed in our analysis, is that features of the base-year interview experience are comparatively strong predictors of attrition, especially interviewer ratings of respondents' comprehension of questions and level of cooperation for the interview.

In this section, we first offer pooled models that combine all three panels, with the justification that the study design was fixed for all three panels. We then offer panel-specific models in order to demonstrate panel-specific variability in the estimates (and to discuss the costs and benefits of pooling). The code released with this report (see the repository, gss-treble-panel-weights, on GitHub; URL here) provides full details on the underlying models: all coding decisions for the predictors, procedures for systematic and multiple imputation of item-specific missing data, a wider range of estimated models than we summarize in the text of this report, and more detail on all models estimated. The tables presented below represent only a small portion of the models that needed to be fit in order to construct the weights.

Pooled Models

We first present estimated logit models of attrition for two overlapping samples that pool all three panels – all base-year panel respondents (N = 6,067) and then all base-year panel respondents who did not transition to ineligibility for either follow-up wave (N = 5,732). Table 3 presents average probability differences, calculated from these estimated logit models, for two types of specifications. First, in columns 1 and 3, we calculate average probability differences from 21 separate logit models that predict attrition based on each predictor variable separately. Second, in columns 2 and 4, we then calculate net average probability differences based on models that specify all 21 predictor variables simultaneously.

---

[6] The West South Central census region includes Texas, Oklahoma, Louisiana, and Arkansas. In addition, she was coded as "89 or older" in the base year, leaving open the possibility that she completed interviews beyond age 100.

Table 3. Average differences in the estimated probability of attrition from underlying bivariate and multiple logistic regression models (with all three panels pooled and for both the base-year eligible sample and the always-in-scope, panel-eligible sample)

Predictor	All base-year respondents (N = 6,067)		Base-year respondents who remained eligible (N = 5,732)	
	Bivariate	All specified simultaneously	Bivariate	All specified simultaneously
Age (orthogonal polynomial)				
Linear (standardized)	0.016	0.029	-0.007	0.007
	(0.007)	(0.009)	(0.007)	(0.009)
Squared (standardized)	0.057	0.042	0.035	0.022
	(0.007)	(0.008)	(0.007)	(0.008)
Cubed (standardized)	0.033	0.025	0.027	0.018
	(0.007)	(0.007)	(0.007)	(0.007)
Female	-0.036	-0.027	-0.020	-0.009
	(0.015)	(0.014)	(0.015)	(0.015)
Race (ref: white)				
Black	0.013	-0.035	0.001	-0.046
	(0.021)	(0.022)	(0.022)	(0.022)
Hispanic	0.137	-0.015	0.135	-0.016
	(0.024)	(0.031)	(0.024)	(0.031)
Asian	0.079	-0.020	0.062	-0.035
	(0.042)	(0.042)	(0.042)	(0.040)
Other	0.201	0.107	0.171	0.091
	(0.078)	(0.068)	(0.081)	(0.071)
Education (ref: high school diploma)				
Less than high school	0.158	0.048	0.132	0.044
	(0.022)	(0.023)	(0.023)	(0.024)
Junior college	-0.011	0.007	0.001	0.010
	(0.027)	(0.027)	(0.027)	(0.027)
Bachelor's degree	-0.014	0.006	-0.011	0.002
	(0.020)	(0.021)	(0.020)	(0.021)
Graduate degree	-0.045	-0.018	-0.039	-0.022
	(0.025)	(0.028)	(0.025)	(0.028)
Household income (log)	-0.028	0.017	-0.016	0.023
	(0.007)	(0.009)	(0.007)	(0.009)
Respondent's dwelling (ref: owns/mortgages home)				
Pays rent	0.071	0.033	0.072	0.039
	(0.016)	(0.020)	(0.016)	(0.020)
Other	0.140	0.132	0.145	0.142
	(0.070)	(0.073)	(0.074)	(0.076)

Table 3 (Continued)

Predictor	All base-year respondents (N = 6,067)		Base-year respondents who remained eligible (N = 5,732)	
	Bivariate	All specified simultaneously	Bivariate	All specified simultaneously
Marital status (ref: married)				
Widowed	0.148	0.073	0.062	0.046
	(0.029)	(0.032)	(0.030)	(0.033)
Divorced	-0.014	0.002	-0.034	-0.019
	(0.021)	(0.021)	(0.021)	(0.021)
Separated	0.093	0.048	0.087	0.043
	(0.047)	(0.045)	(0.049)	(0.045)
Never married	0.046	0.037	0.035	0.022
	(0.018)	(0.021)	(0.018)	(0.021)
Citizenship status (ref: natural born citizen)				
Naturalized citizen	0.032	0.013	0.018	-0.002
	(0.029)	(0.032)	(0.029)	(0.032)
Non-citizen	0.320	0.260	0.314	0.245
	(0.028)	(0.038)	(0.030)	(0.041)
Region (ref: New England)				
Middle Atlantic	-0.012	-0.003	-0.013	0.002
	(0.043)	(0.041)	(0.043)	(0.041)
East North Central	-0.040	-0.014	-0.039	-0.009
	(0.041)	(0.040)	(0.042)	(0.040)
West North Central	-0.049	-0.003	-0.050	-0.003
	(0.047)	(0.047)	(0.047)	(0.047)
South Atlantic	0.012	0.021	0.013	0.027
	(0.041)	(0.039)	(0.041)	(0.039)
East South Central	-0.014	0.016	-0.005	0.025
	(0.048)	(0.047)	(0.048)	(0.047)
West South Central	0.015	0.020	0.009	0.020
	(0.044)	(0.043)	(0.044)	(0.044)
Mountain	-0.039	-0.015	-0.047	-0.018
	(0.045)	(0.044)	(0.046)	(0.044)
Pacific	0.026	0.019	0.037	0.028
	(0.042)	(0.041)	(0.043)	(0.041)
Vocabulary score (number correct out of 10)	-0.038	-0.022	-0.033	-0.021
	(0.004)	(0.005)	(0.004)	(0.005)
Interviewer's age	-0.002	-0.001	-0.002	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Female interviewer	-0.036	-0.027	-0.040	-0.033
	(0.019)	(0.020)	(0.020)	(0.020)

Table 3 (Continued)

Predictor	All base-year respondents (N = 6,067)		Base-year respondents who remained eligible (N = 5,732)	
	Bivariate	All specified simultaneously	Bivariate	All specified simultaneously
Interviewer's race (ref: white)				
Black	-0.000	0.007	-0.006	0.010
	(0.021)	(0.022)	(0.021)	(0.023)
Hispanic	0.095	-0.039	0.101	-0.026
	(0.027)	(0.031)	(0.028)	(0.031)
Asian	0.183	0.141	0.181	0.142
	(0.059)	(0.056)	(0.061)	(0.058)
Two or more races	-0.014	-0.028	-0.014	-0.019
	(0.042)	(0.041)	(0.042)	(0.042)
Interviewer years of experience	-0.004	-0.003	-0.003	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)
Length of interview (in hours)	-0.002	-0.005	-0.012	-0.009
	(0.012)	(0.012)	(0.012)	(0.012)
Days elapsed since first interview of the year and respondent's interview (divided by 10)	-0.001	0.002	0.001	0.003
	(0.002)	(0.002)	(0.002)	(0.002)
Interview mode (ref: in person)				
Over the phone	0.020	0.031	0.037	0.038
	(0.024)	(0.026)	(0.025)	(0.026)
Incentive for base-year (ref: monetary incentive)				
Non-monetary incentive	0.050	0.061	0.030	0.053
	(0.063)	(0.061)	(0.061)	(0.061)
No incentive	0.024	0.042	0.014	0.045
	(0.014)	(0.017)	(0.015)	(0.018)
Respondent's attitude toward the interview (ref: friendly, interested)				
Cooperative	0.076	0.047	0.064	0.039
	(0.020)	(0.020)	(0.021)	(0.021)
Restless, impatient	0.168	0.134	0.173	0.144
	(0.038)	(0.037)	(0.039)	(0.038)
Hostile	0.485	0.445	0.464	0.433
	(0.094)	(0.104)	(0.118)	(0.121)

Table 3 (Continued)

Predictor	All base-year respondents (N = 6,067)		Base-year respondents who remained eligible (N = 5,732)	
	Bivariate	All specified simultaneously	Bivariate	All specified simultaneously
Respondent's understanding of the questions (ref: good)				
Fair	0.131	0.030	0.116	0.032
	(0.022)	(0.023)	(0.023)	(0.023)
Poor	0.243	0.036	0.202	0.034
	(0.050)	(0.045)	(0.054)	(0.047)
Language of interview (ref: English)				
Spanish	0.256	0.043	0.251	0.031
	(0.034)	(0.051)	(0.036)	(0.051)

Notes: The models are weighted with the base-year cross-sectional weight, *wtssnr*, scaled to adjust for slight differences in panel sample sizes. Item-specific missing values in the base year for predictors that could not be determined systematically from follow-up waves were imputed with random-forest imputation. The models presented in the table are for the models fit with the first set of these imputations (but the estimated probabilities used for weight construction are based on all six datasets with imputations, suitably estimated with multiple imputation methods). Because the imputed values are selected by random forests, they are less variable than typically appears to be the case for parametric multiple-imputation methods. We used random-forest estimation because our judgment is that, for the GSS with its relatively small amount of item-specific missing data, random forests yield more reasonable imputations (in particular, imputations that were less likely to be at the minimum and maximum values of the observed ranges of each variable, which we interpret as evidence of overfitting). Finally, only the variables for owning or renting one's dwelling and the vocabulary test had substantial missing data, but this missing data is mostly from randomly assigned ballots that did not offer the relevant questions to all respondents.

For reference to code and results:
> Column 1 is drawn from column D, column Q, and similar columns in docs/attr-weight-bivariate-all.xls. The corresponding logit coefficients are in columns C, P, etc.
> Column 2 is column D in docs/attr-weight-multivariate-pooled-all.xls. The corresponding logit coefficients are in column C, and the pooled logit coefficients are in column B.
> Column 3 and 4 are in analogous positions in docs/attr-weight-bivariate-all.xls and docs/attr-weight-multivariate-pooled-insc.xls.

Overall, attrition is not as strongly patterned as one might expect (or fear). A simple summary of the patterning of attrition can be calculated as a difference in predictive classification: the difference between (a) the estimated probability of attrition among those who actually did attrite and (b) the estimated probability of attrition among those who did not attrite. If the models are strongly predictive, these average probability estimates should be very different. For the model in column 2, the estimated average probabilities are 0.406 and 0.331, respectively. For the model in column 4, the analogous values are 0.367 and 0.306.

Nonetheless, the models suggest patterns that are sensible and important, as can be seen by differences in estimated average probabilities within the columns of Table 3 and across the models within the rows of Tables 3. Some important patterns are:

- The predicted attrition pattern is slightly stronger for overall attrition (columns 1 and 2) than for attrition within the in-scope subset of the base-year sample (columns 3 and 4). Although the general pattern can be seen for the polynomial specification of age in the first three rows, the widowhood difference is the easiest to interpret. The average probability difference in attrition for widows and widowers is higher, in comparison to those currently married. For the full base-year sample, the unadjusted probability difference is 0.148, suggesting that 14.8 percent more widows and widowers attrited than those who were never married. When adjusted in a model with all 21 predictors, the adjusted probability difference declines by half to 0.073. However, when the sample is restricted to those who were not known to have transitioned to ineligibility (typically, by dying), the average probability differences fall to 0.062 and 0.046, respectively. Thus, widowhood is a stronger predictor for the full base-year sample, rather than the always-in-scope sample, because death is more common among widows and widowers and also a transition mechanism from GSS eligibility to GSS ineligibility.

- Demographic characteristics as well as measures of socio-economic standing predict attrition to some degree, but their net associations are typically much smaller than their unadjusted associations (i.e., as shown in comparisons of columns 1 and 3 to columns 2 and 4).

- Base-year respondents who were not citizens were substantially more likely to attrite (32.0 and 31.4 percent for each sample), even after adjustments for other predictors (26.0 and 24.5 percent for each sample, respectively).[7]

- The 10-item vocabulary test remains predictive, even after adjustments for other predictors. As shown in columns 2 and 4, the probability of attrition was higher by 8.8

[7] We should also clarify that our definition of "non-citizen" is strict, based on all observed information. In particular, respondents who were not citizens for the base year but who become citizens by the time they were reinterviewed for one or both follow-up waves were excluded from the group of non-citizens that defined the variable for the logit models (under the assumption that such respondents were in the process of acquiring citizenship at the time of their base-year interview).

percent and 8.4 percent between those who differed by 4 correct answers on the 10-item test.

- Although interviewer characteristics are not strongly predictive, the interviewers' ratings of respondents' behavior during the base-year interview are among the strongest predictors of attrition. In comparison to those who were rated as "friendly, interested," respondents who were rated as only "cooperative," "restless, impatient" and "hostile" were increasingly less likely to agree to a follow-up interview. These associations are diminished only to a small degree in the full models in columns 2 and 4, when other predictors are specified as well. Somewhat similarly, respondents who were rated as having only a fair or poor understanding of the questions, as opposed to a good understanding, were more likely to attrite. This second set of associations, however, declined substantially for the full models, suggesting that education and the vocabulary test, for example, share some predictive variance with comprehension of the questions.

Many more fine-grained details of the attrition patterns are revealed by Table 3, as well as by the additional models offered with the distributed code and results.

Finally, we should explain two features of the model specification. First, the bivariate average probability differences presented in columns 1 and 3, based on 42 separate logit models, are offered only to aid interpretations. For the weights we construct, we will draw estimated probabilities of attrition from full models that specify all predictors, as in columns 2 and 4. The bivariate average probability differences in columns 1 and 3, nonetheless, help to make clear that the average probability differences from net associations can be hard to interpret on their own. For example, the average probability difference for taking the base-year interview in Spanish rather than English, and for identifying as Hispanic rather than non-Hispanic white, are both substantial and positive only for the bivariate models. The full models that adjust for many other factors at the same time do not eliminate these bivariate associations but rather suggest that these predictors share most of their predictive covariance with other measured characteristics of respondents. It is not the case that the estimated partial associations for the full models are more "correct," only that the full models are collectively more predictive and can be expected to generate estimated probabilities for individuals that are more informative.

Second, we did not push the full models to more highly parameterized forms by specifying cross-product interactions between predictors. A more perfect specification would surely require a parameterization that allowed for at least some such conditional associations, and it is possible that additional modeling could reveal them effectively. Our goal in this report is to develop weights from a basic set of models that others may wish to build upon. That being said, we consider these specifications suitable enough, by our interpretation of the standards in the literature on panel weighting, for developing weights that can improve subsequent empirical analysis.

Three Panel-Specific Models

We next report results from estimated logit models separately for each of the three panels in order to assess the extent to which the pooled models presented in Table 3 may be hiding important panel-specific heterogeneity in attrition patterns. Although the design of the GSS was fixed, and the prescribed field procedures stable across all years, it is possible that changes between 2008 and 2014 in the experiences of the respondents altered the patterns of attrition, and that these could have interacted with variation in how field procedures were carried out.

Table 4 presents average probability differences for each of the three panels. For brevity, we offer only the models for the smaller always-in-scope realized sample (i.e., not the larger base-year realized sample) and the full specifications that use all 21 predictors simultaneously (i.e., not the separate bivariate models for each predictor).[8] To facilitate comparisons, we have pasted the analogous values from the relevant pooled model into the first column of Table 4 (i.e., exactly the same values that were presented in column 4 of Table 3). The panel-specific estimated probability differences are then presented in columns 2 through 4 of Table 4.

As noted above, net associations such as those in Table 4 are hard to interpret because predictive covariance is shared by subsets of the predictors. It is unclear whether the variability across panels has been amplified by either overfitting or the lumpiness of individual characteristics, the risk of which both increase when the panels are analyzed separately. An alternative, and potentially clearer, basis for between-panel comparisons would be to compare the distributions of estimated probabilities of attrition from each model, as we present next.

---

[8] A wider range of models is available in the shared code and results that we offer, and they are consistent with what is presented in Table 4.

Table 4. Average differences in the estimated probability of attrition from underlying multiple logistic regression models with all panels pooled and with each panel separately (limited to the always-in-scope, panel-eligible sample)

Predictor	All pooled (N = 5,732)	2006 panel (N = 1,891)	2008 panel (N = 1,918)	2010 panel (N = 1,923)
Age (orthogonal polynomial)				
Linear (standardized)	0.007	-0.004	-0.005	0.025
	(0.009)	(0.015)	(0.016)	(0.016)
Squared (standardized)	0.022	0.016	0.018	0.040
	(0.008)	(0.013)	(0.014)	(0.014)
Cubed (standardized)	0.018	0.035	0.012	0.011
	(0.007)	(0.012)	(0.013)	(0.012)
Female	-0.009	-0.005	-0.044	0.028
	(0.015)	(0.025)	(0.025)	(0.024)
Race (ref: white)				
Black	-0.046	-0.045	0.008	-0.128
	(0.022)	(0.038)	(0.041)	(0.031)
Hispanic	-0.016	-0.019	0.055	-0.079
	(0.031)	(0.052)	(0.059)	(0.045)
Asian	-0.035	-0.036	0.013	-0.052
	(0.040)	(0.068)	(0.077)	(0.062)
Other	0.091	-0.006	0.254	0.057
	(0.071)	(0.127)	(0.105)	(0.111)
Education (ref: high school diploma)				
Less than high school	0.044	0.038	0.037	0.072
	(0.024)	(0.042)	(0.043)	(0.040)
Junior college	0.010	-0.001	-0.014	0.041
	(0.027)	(0.044)	(0.046)	(0.049)
Bachelor's degree	0.002	-0.012	0.001	0.005
	(0.021)	(0.036)	(0.038)	(0.034)
Graduate degree	-0.022	-0.060	-0.034	0.021
	(0.028)	(0.047)	(0.048)	(0.046)
Household income (log)	0.023	0.035	0.049	-0.007
	(0.009)	(0.016)	(0.018)	(0.014)
Respondent's dwelling (ref: owns/mortgages home)				
Pays rent	0.039	0.060	0.024	0.031
	(0.020)	(0.033)	(0.038)	(0.032)
Other	0.142	0.259	0.159	-0.091
	(0.076)	(0.121)	(0.124)	(0.094)

Table 4 (Continued)

Predictor	All pooled (N = 5,732)	2006 panel (N = 1,891)	2008 panel (N = 1,918)	2010 panel (N = 1,923)
Marital status (ref: married)				
Widowed	0.046	0.122	0.072	-0.035
	(0.033)	(0.057)	(0.060)	(0.052)
Divorced	-0.019	0.007	-0.040	-0.033
	(0.021)	(0.034)	(0.039)	(0.036)
Separated	0.043	0.118	0.025	-0.015
	(0.045)	(0.079)	(0.076)	(0.065)
Never married	0.022	0.056	0.020	0.013
	(0.021)	(0.036)	(0.037)	(0.036)
Citizenship status (ref: natural born citizen)				
Naturalized citizen	-0.002	-0.150	0.061	0.065
	(0.032)	(0.045)	(0.055)	(0.054)
Non-citizen	0.245	0.219	0.139	0.336
	(0.041)	(0.067)	(0.071)	(0.068)
Region (ref: New England)				
Middle Atlantic	0.002	-0.013	0.020	-0.045
	(0.041)	(0.074)	(0.074)	(0.064)
East North Central	-0.009	-0.038	-0.026	0.014
	(0.040)	(0.072)	(0.072)	(0.064)
West North Central	-0.003	-0.018	-0.036	0.049
	(0.047)	(0.081)	(0.085)	(0.078)
South Atlantic	0.027	0.014	0.008	0.039
	(0.039)	(0.071)	(0.069)	(0.063)
East South Central	0.025	-0.003	-0.011	0.069
	(0.047)	(0.083)	(0.086)	(0.074)
West South Central	0.020	0.043	-0.056	0.030
	(0.044)	(0.077)	(0.080)	(0.071)
Mountain	-0.018	0.012	-0.113	0.019
	(0.044)	(0.078)	(0.078)	(0.073)
Pacific	0.028	0.062	-0.131	0.091
	(0.041)	(0.074)	(0.072)	(0.066)
Vocabulary score (number correct out of 10)	-0.021	-0.026	-0.014	-0.016
	(0.005)	(0.010)	(0.010)	(0.008)
Interviewer's age	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Female interviewer	-0.033	-0.056	0.047	-0.077
	(0.020)	(0.034)	(0.036)	(0.033)

Table 4 (Continued)

Predictor	All pooled (N = 5,732)	2006 panel (N = 1,891)	2008 panel (N = 1,918)	2010 panel (N = 1,923)
Interviewer's race (ref: white)				
Black	0.010	0.012	-0.005	0.071
	(0.023)	(0.041)	(0.040)	(0.040)
Hispanic	-0.026	-0.042	0.008	-0.004
	(0.031)	(0.051)	(0.059)	(0.049)
Asian	0.142	0.070	0.407	0.257
	(0.058)	(0.071)	(0.159)	(0.120)
Two or more races	-0.019	--	0.026	-0.012
	(0.042)	--	(0.083)	(0.053)
Interviewer years of experience	-0.002	-0.003	-0.001	-0.014
	(0.002)	(0.004)	(0.003)	(0.009)
Length of interview (in hours)	-0.009	0.002	-0.009	-0.037
	(0.012)	(0.024)	(0.018)	(0.025)
Days elapsed since first interview of the year and respondent's interview (divided by 10)	0.003	0.002	0.009	0.003
	(0.002)	(0.004)	(0.004)	(0.004)
Interview mode (ref: in person)				
Over the phone	0.038	-0.005	0.057	0.072
	(0.026)	(0.050)	(0.043)	(0.044)
Incentive for base-year (ref: monetary incentive)				
Non-monetary incentive	0.053	0.045	0.016	0.146
	(0.061)	(0.087)	(0.119)	(0.127)
No incentive	0.045	0.024	0.102	0.008
	(0.018)	(0.033)	(0.031)	(0.030)
Respondent's attitude toward the interview (ref: friendly, interested)				
Cooperative	0.039	0.045	0.047	0.021
	(0.021)	(0.037)	(0.035)	(0.034)
Restless, impatient	0.144	0.215	0.090	0.115
	(0.038)	(0.066)	(0.064)	(0.063)
Hostile	0.433	0.422	0.506	0.570
	(0.121)	(0.176)	(0.158)	(0.111)

Table 4 (Continued)

Predictor	All pooled (N = 5,732)	2006 panel (N = 1,891)	2008 panel (N = 1,918)	2010 panel (N = 1,923)
Respondent's understanding of the questions (ref: good)				
Fair	0.032	0.022	0.027	0.057
	(0.023)	(0.044)	(0.043)	(0.036)
Poor	0.034	-0.142	0.193	0.075
	(0.047)	(0.055)	(0.084)	(0.077)
Language of interview (ref: English)				
Spanish	0.031	0.096	0.036	-0.000
	(0.051)	(0.088)	(0.092)	(0.088)

Notes: See notes for Table 3, which apply to this table as well.

For reference to code and results:

> Column 1 is the same as column 4 from Table 3. See notes above.

> Column 2, 3, and 4 are column D in docs/attr-weight-multivariate-p2006-insc.xls, docs/attr-weight-multivariate-p2008-insc.xls, and docs/attr-weight-multivariate-p2010-insc.xls, respectively. The corresponding logit coefficients are in column C.

Graphical Representations of the Distributions of the Estimated Probabilities

Figures 1 and 2 present scatterplots of predicted probabilities from four models specified like those in Tables 3 and 4 above. For Figure 1, the probabilities are drawn from models estimated for the full base-year sample. For Figure 2, the probabilities are drawn from models estimated for the always-in-scope sample.[9]

Raw estimated probabilities for each individual. Consider first the three scatterplots in the left column of each figure. These three scatterplots have the same horizontal dimension: the estimated probabilities of each individual from the pooled model. The vertical dimensions, in contrast, differ across the three scatterplots. They are the estimated probabilities of each individual from each panel-specific model.

As shown in these scatterplots, the probabilities are strongly related, with similar upward sloping clouds of points for each panel-by-pooled set of estimated probabilities. However, the clouds are not close to lines, as would be the case if the pooled model was equivalent to each panel-specific model.

Is the vertical spread of the estimated probabilities from the panel-specific models (on either side of the estimated probabilities from the pooled model) the result of genuine improvement in the modeling of attrition for each panel? Or, instead, is the vertical spread random variation attributable to the smaller panel-specific samples? Or, most worrying, is the spread the result of having overfit the panel-specific models to the data, asking too much of the data by trying to estimate 21 net effects simultaneously? There is no clear answer, but we are inclined to believe that overfitting is a genuine concern (but not so convinced it is present that the panel-specific models should be entirely abandoned). This brings us to a potential solution depicted in the right column of each figure, explained below.

---

[9] For Figure 1, we limit the scatterplots to non-attriters in order to show the distributions of the probabilities that are used to construct the weights (since the weights are constructed only for non-attriters). Scatterplots that include the predicted probabilities of attriters are qualitatively similar, with additional probabilities disproportionately in the right tails of the distributions, just as one would anticipate.
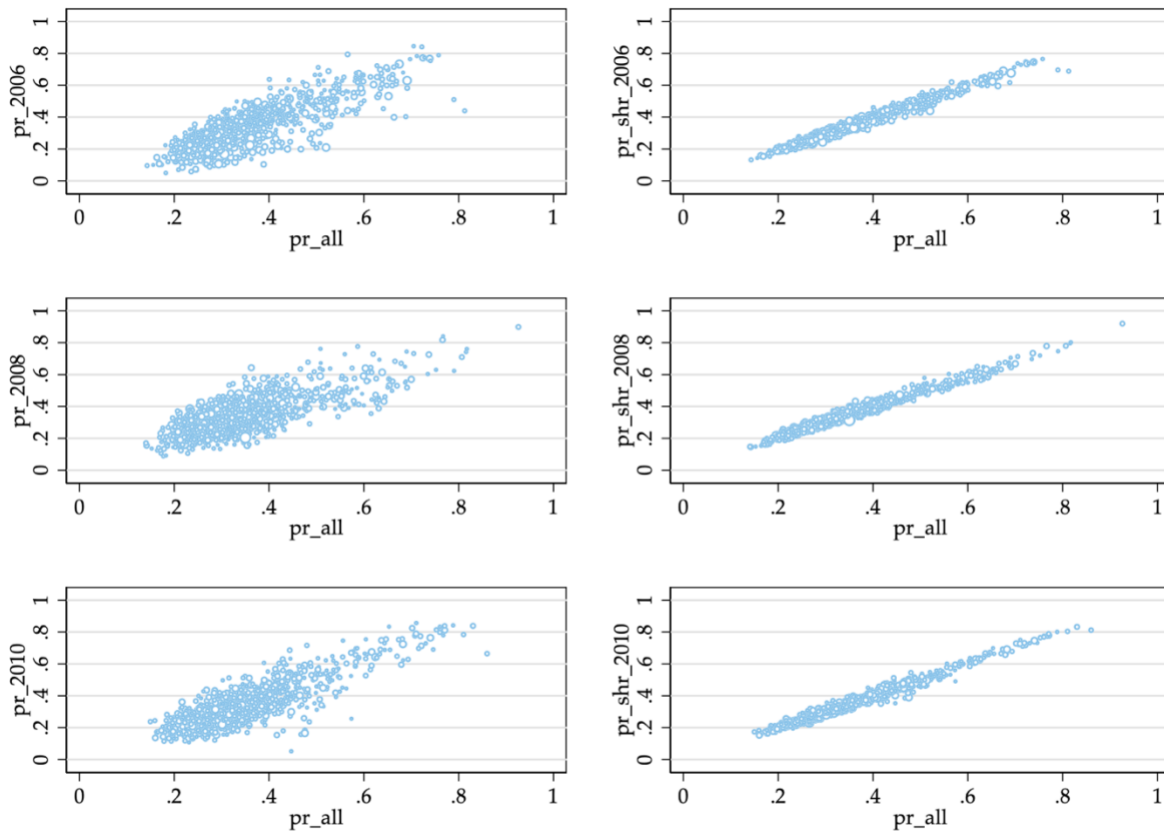
Figure 1. Scatterplots for the base-year eligible sample of year-specific estimated probabilities of attrition (left column, as pr_2006, pr_2008, and pr_2010) and shrunken year-specific estimated probabilities of attrition (right column, as pr_shr_2006, pr_shr_2008, and (pr_shr_2010) by estimated probabilities of attrition from the model that pools all three panels

Notes: Size of marker is proportional to the baseline weight, *wtpannr123*.
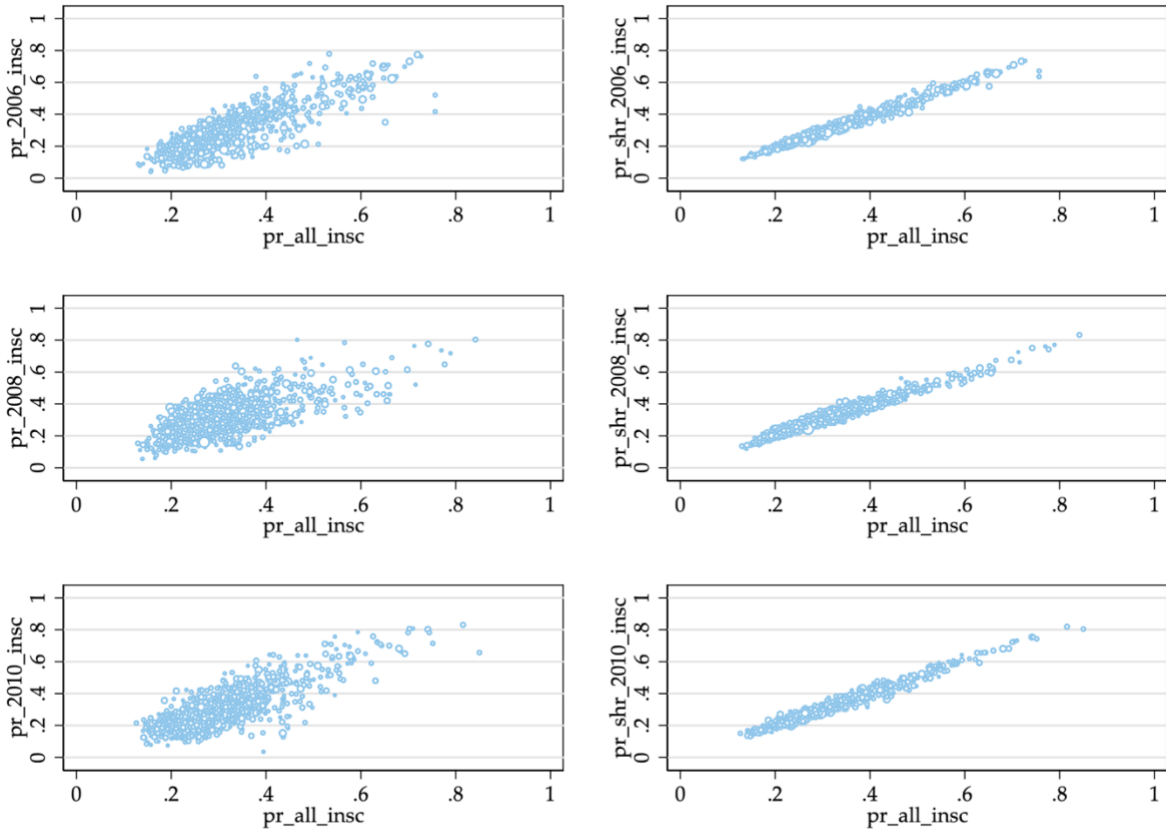
Figure 2. Analogous scatterplots to those in Figure 1 but estimated for the always-in-scope, panel eligible sample

Notes: Size of marker is proportional to the baseline weight, *wtpannr123*.

Shrunken panel-specific estimated probabilities of attrition. How should we interpret the vertical dispersion around the pooled-model probabilities within the three scatterplots in the left columns of both Figures 1 and 2? As briefly discussed above, we do not have a clear answer. This uncertainty suggests the value of what is sometimes labeled an "empirical Bayes" approach. In the right column of Figures 1 and 2, we offer three additional scatterplots, where the horizontal dimension is the same as in the left column. However, the vertical dimension is now a set of "shrunken" estimated probabilities for each individual.

In particular, these shrunken estimated probabilities are precision-weighted averages of the estimated probabilities from the pooled model and the respective panel-specific model, where precision is defined as the inverse of the variance of each individual's linear prediction from each of the two models utilized. For example, the shrunken probabilities for 2006 are calculated as

$$\widehat{p}_{i\,shrunken,2006} = \frac{1}{c_i}\left(\frac{1}{Var(\boldsymbol{x}_i'\widehat{\boldsymbol{b}}_{pooled})}\widehat{p}_{i\,pooled} + \frac{1}{Var(\boldsymbol{x}_i'\widehat{\boldsymbol{b}}_{2006})}\widehat{p}_{i\,2006}\right)$$

where the constant c_i is the sum of the precision weights

$$c_i = \frac{1}{Var(\boldsymbol{x}_i'\widehat{\boldsymbol{b}}_{pooled})} + \frac{1}{Var(\boldsymbol{x}_i'\widehat{\boldsymbol{b}}_{2006})}$$

that scales the precision-weighted probabilities upward to the unit interval. In addition, the variance terms are estimated quantities from the data that vary over individuals *i* because of the variation across the individual-specific row vectors $\boldsymbol{x}_i'$.

Table 5 presents descriptive statistics for the raw panel-specific probabilities and their shrunken variants. The two panels correspond to Figures 1 and 2. The shrinkage transformation reduces the variance of the panel-specific estimated probabilities (and usually pulls in extreme values as a byproduct).

However, there is some variation across years, necessitating "usually" in the parentheses of the paragraph above. If one believes, as we do, that the true attrition process does not vary much across panels, then the estimated probabilities from the pooled model are based on more systematic information (i.e., are less likely to have variance attributable to chance or overfitting).

This is the explanation for what would appear to be an anomaly in the maximum estimated probabilities for 2008. Consider the first panel of Table 5. The maximum value for the pooled probability is 0.926 (and this value is for the elderly, non-citizen, Spanish speaker described above in the discussion of attrition due to ineligibility). The pooled model gives her a higher estimated value than does the 2008 model, which suggests a value of only 0.899. Her shrunken estimated probability is then higher at 0.920. In other words, her shrunken probability is more extreme on an absolute scale than is her panel-specific value. While this may seem like an anomaly, it suggests an overall point that is important and can be explained

graphically. The shrinkage only affects the dispersion of points in the vertical dimension of the figures, since the transformation does not alter the dispersion along the horizontal dimension. Nonetheless, the shrinkage can pull points across the diagonal (i.e., an implicit 1-to-1 identity line), and this is what occurs for the seeming anomaly (moving upwards from 0.899, crossing 0.9, and becoming 0.920). Overall, the shrinkage yields estimated probabilities in the right columns of Figures 1 and 2 that are closer in alignment with the pooled probabilities, and hence the diagonal. But, there is no requirement that the shrinkage prevent movement across the diagonal. And, because that is possible, the "usually" above is necessary.

Table 5. Descriptive statistics of estimated probabilities of attrition from a pooled model, from models for each panel separately, and then for each panel separately but shrunken toward the estimates from the pooled model

	N	Mean	SD	Min	Max
Based on the base-year eligible sample:					
Probabilities (all years pooled)	3,875	0.334	0.107	0.140	0.926
Probabilities in 2006	1,276	0.313	0.133	0.050	0.846
Shrunk to pooled estimates	1,276	0.333	0.112	0.132	0.765
Probabilities in 2008	1,295	0.340	0.109	0.088	0.899
Shrunk to pooled estimates	1,295	0.333	0.099	0.146	0.920
Probabilities in 2010	1,304	0.323	0.129	0.052	0.857
Shrunk to pooled estimates	1,304	0.330	0.107	0.152	0.832
Based on the sample that is in scope all waves:					
Probabilities (all years pooled)	3,875	0.308	0.097	0.126	0.850
Probabilities in 2006	1,276	0.286	0.125	0.040	0.779
Shrunk to pooled estimates	1,276	0.308	0.104	0.121	0.735
Probabilities in 2008	1,295	0.318	0.106	0.056	0.804
Shrunk to pooled estimates	1,295	0.308	0.090	0.120	0.833
Probabilities in 2010	1,304	0.295	0.121	0.036	0.831
Shrunk to pooled estimates	1,304	0.302	0.097	0.131	0.819

Notes: Data are weighted by *wtpannr123*.

Altogether, we have estimated six sets of estimated probabilities. For two overlapping samples (all base-year respondents and the subsample who remained in scope for all waves), we have produced pooled $\hat{p}_l$, panel-specific $\hat{p}_l$, and shrunken panel-specific $\hat{p}_l$. We will use these six sets of estimated probabilities to construct six weights. While all weights can be used, we explain below that we (weakly) favor the weights that use the shrunken probabilities estimated from the always-in-scope sample.

WEIGHT CONSTRUCTION

Many books on procedures to construct survey weights are available (e.g., Levy and Lemeshow 2008; Thompson 2012; Valliant, Dever, and Kreuter 2013), and we will not provide a full explanation of the rationale for attrition weighting in this section. We assume that readers who have gotten this far into this report already know well the value and limitations of panel weighting approaches.[10] We first explain the components of the cross-sectional and panel weights constructed by NORC, and we then detail the six panel weights that we have constructed to extend the NORC panel weight to also account for attrition.

NORC's Baseline Weights

For the 2006, 2008, and 2010 cross-sectional samples, NORC's recommended cross-sectional weight is *wtssnr*. This weight is based on the sampling design of the GSS, and it is necessary to first explain its components. After presenting these details, we then explain NORC's own panel weight, *wtpannr123*, which is based on *wtssnr* and is the weight that we modify for the six panel weights developed for this report.

    NORC's Cross-Sectional Weight (wtssnr). The GSS sample design is a complex stratified multi-stage sample, but it is nonetheless "equal probability" at the household level.[11] As a result, all households in the target population have an equal probability of inclusion, and the GSS would be self-weighting at the household level if household-level response rates were uniform.

    The GSS design specifies that only one adult respondent be selected for interview from each participating household. Thus, the GSS is not self-weighting for the adult target population because the number of adults in each household varies across households. The weight variable *wtssnr* scales up each household's selected respondent by a factor that is proportional to the number of adults in the household, as coded in the GSS variable *adults*.[12] This scaling for the household count of adults represents about two-thirds of the variance of *wtssnr*.

    In addition, *wtssnr* adjusts for two additional factors for the years relevant for the panel. After an initial field period for each cross-sectional sample in 2006, 2008, and 2010, non-responding households were randomly subsampled and assigned for additional effort to gain cooperation. The goal of this subsampling was to focus targeted effort on a smaller number of

---

[10] In addition, while we will provide many of the details in this section of how we constructed the weights that we offer, yet more details are revealed in the heavily commented code that we offer along with this report. These additional details include matters such as how we scaled the weights at various points in their construction so that they summed to the appropriate values.

[11] For details of the sampling design, see Appendix A of Smith, Davern, Freese, and Morgan (2019). See also Harter, Eckman, English, and O'Muircheartaigh (2010).

[12] The number of adults is determined by household enumeration at the point of first contact with the household, and the respondent is selected using a Kish allocation that is (nearly equivalent) to a random draw (see Kish 1965, section 11.3).

reluctant households, under the assumption that most households would eventually submit to NORC's charm if enough effort were expended. The weight *wtssnr* scales up households subsampled for intensive effort because these households must represent those subsampled and those not subsampled. Households that were selected for additional effort are indicated by the GSS variable *phase*.

Finally, *wtssnr* incorporates a simple design-based non-response adjustment. For each geographic area chosen according to NORC's national sampling design, the household-level response rate can be calculated after data collection is complete. Response rates vary modestly across areas for unknown reasons, possibly due to chance, possibly due to non-randomly distributed skill across interviewers in gaining cooperation, and so forth. With these area-level response rates, households who participated in each area can be scaled up or down by a factor that is inversely proportional to the calculated household-level response rate in each area. This adjustment factor does not use any information on households, nor external information known or estimated for each sampling area.[13] It is simply an adjustment based on variation in the realized raw response rates across areas.

NORC's Panel Weight (*wtpannr123*). With the panel dataset, NORC distributes a scaled version of *wtssnr*, which is labeled *wtpannr123*. The "nr" in the title might suggest that it adjusts for attrition-based panel non-response of the sort that we modeled above. In fact, *wtpannr123* adjusts only for non-response in the cross-section, and only in the mild manner as does *wtssnr* based on variation across sampling areas. The panel weight, *wtpannr123*, is a rescaled version of *wtssnr* that is set to missing for those who did not participate in all three waves of the panel. It is then scaled to sum to the raw number of respondents who participated in all three waves of each corresponding panel.[14]

Panel Weights that Adjust for Attrition

The key idea for the construction of a suitable panel weight that adjusts for attrition is to scale the baseline weights of all respondents who participate in all three waves of the panel by an adjustment factor based on their estimated inverse probability of participating in all three waves of the panel. For the GSS panels, the adjustment factor has the form

$$\frac{1}{1 - \hat{p}_i}$$

---

[13] This non-response adjustment does not attempt to align the realized sample to population estimates by post-stratification to known or well estimated marginals. Indeed, it is a matter of pride for the GSS that post-stratification is not required to bring the realized GSS samples in line with many basic features of the target population. Nonetheless, the GSS does continue to over-represent women, and the geography-based non-response adjustment does not eliminate this mismatch (because this particular response pattern does not vary substantially by geography).

[14] A companion weight, *wtpannr12*, is available for those who participated in the base-year and first follow-up only. Versions are also available without the area-based non-response adjustment, and these are *wtpan12* and *wtpan123*. See Smith and Schapiro (2017) for additional details.

where $\hat{p}_i$ is an estimated probability of attrition, such as those summarized above in Tables 5. Respondents who have observed characteristics most similar to those who attrited are given more weight by this scaling factor. As $\hat{p}_i$ increases, the scaling factor increases because the denominator decreases.

The panel weights we have developed in this report are constructed as

$$\left(\frac{1}{1 - \hat{p}_i}\right) wtpannr123_i$$

where $wtpannr123_i$ is the NORC weight for each panel respondent explained above. The six alternative weights are generated by setting $\hat{p}_i$ equal to each of the six sets of probabilities estimated above: pooled $\hat{p}_i$, panel-specific $\hat{p}_i$, and shrunken panel-specific $\hat{p}_i$, each estimated from two overlapping samples (all base-year respondents and the subsample of those who remained in scope).

Table 6 presents descriptive statistics of six sets of adjustment factors, based on the probabilities presented in Table 5. For brevity, we label the factors as "scaled inverse probabilities of non-attrition" and assign the acronym SIPNA. For interpretability, the ratios $1/(1 - \hat{p}_i)$ are best scaled (very slightly) to ensure that they sum to the appropriate year-specific base-year sample size.[15] With this scaling, the SIPNAs can be interpreted as the number of base-year respondents that each non-attriting panel respondent represents. In this case, because the probability of attrition is, on average, about 0.33 (see Table 5), the multiplier is, on average, typically about 1/(1-0.33) or 1.5. Thus, on average, each panel respondent represents about 1.5 base-year respondents. However, the minima and maxima in the final column of Table 6 show how much the SIPNAs can differ from 1.5.

Table 7 presents descriptive statistics of seven weights, the NORC baseline weight, *wtpannr123*, and the six new weights that adjust for panel attrition based on the estimated probabilities plotted in Figures 1 and 2 and summarized in Table 5. Figure 3 presents the six new weights in Table 7 as the vertical dimensions of six scatterplots where the NORC baseline weight is the common horizontal dimension. (The dark line in each scatterplot is a 1-to-1 diagonal identity line and is for interpretive reference only.) Both Table 7 and Figure 3 show that the attrition adjustment increases the variance of the weights, differentially so based on the estimated attrition probability that is specified. Most of this increase is a consequence of the shift in scale (with the mean moving from 1 to about 1.5). If these weights were to be regarded as base weights for a new sample, we could rescale them to have a mean of 1 and thereby reduce nominally the variance of the weight. We do not do so because these are not base weights, and the scale produced by the inverse probabilities yields a scale with a meaningful interpretation (and because, for nearly all modern software programs, only the relative sizes of weights are used for model estimation when specified correctly).

---

[15] Modern software produces estimates that are invariant to such scaling, if the weight is specified properly (such as an "aweight" or "pweight" in Stata). The scaling is uniform for all respondents in each panel, based on a single constant for each panel, and thus the relative sizes of the weights do not change.

Table 6. Descriptive statistics of scaled inverse probabilities of non-attrition (SIPNAs) based on the estimated probabilities from Table 5

	N	Mean	SD	Min	Max
Based on the base-year eligible sample:					
SIPNAs (all years pooled)	3,875	1.566	0.459	1.158	13.610
SIPNAs in 2006	1,276	1.567	0.488	1.072	6.608
Shrunk to pooled estimates	1,276	1.567	0.396	1.155	4.266
SIPNAs in 2008	1,295	1.562	0.442	1.085	9.821
Shrunk to pooled estimates	1,295	1.562	0.474	1.177	12.547
SIPNAs in 2010	1,304	1.567	0.572	1.051	6.941
Shrunk to pooled estimates	1,304	1.567	0.453	1.187	6.009
Based on the sample that is in scope all waves:					
SIPNAs (all years pooled)	3,875	1.479	0.335	1.130	6.651
SIPNAs in 2006	1,276	1.482	0.398	1.053	4.582
Shrunk to pooled estimates	1,276	1.482	0.327	1.129	3.748
SIPNAs in 2008	1,295	1.481	0.328	1.037	4.991
Shrunk to pooled estimates	1,295	1.481	0.305	1.136	5.984
SIPNAs in 2010	1,304	1.475	0.439	1.029	5.857
Shrunk to pooled estimates	1,304	1.475	0.343	1.150	5.532

Notes: The acronym SIPNA is "scaled inverse probability of non-attrition."

Table 7. Descriptive statistics of NORC's baseline panel weight and six additional weights that also adjust for attrition

| Name	Description	N	Mean	SD	Min	Max
wtpannr123	NORC baseline panel weight	3,875	1.000	0.627	0.307	6.323
Based on the base-year eligible sample:						
wt_no_pooling	Estimated separately by panel	3,875	1.566	1.127	0.380	13.536
wt_pooled	Estimated with panels pooled	3,875	1.566	1.110	0.397	14.864
wt_shrunken	Shrunk to pooled estimates	3,875	1.566	1.102	0.394	13.703
Based on the sample that is in scope all waves:						
wt_no_pooling_insc	Estimated separately by panel	3,875	1.479	1.050	0.359	11.846
wt_pooled_insc	Estimated with panels pooled	3,875	1.479	1.029	0.379	10.990
wt_shrunken_insc	Shrunk to pooled estimates	3,875	1.479	1.025	0.375	10.778
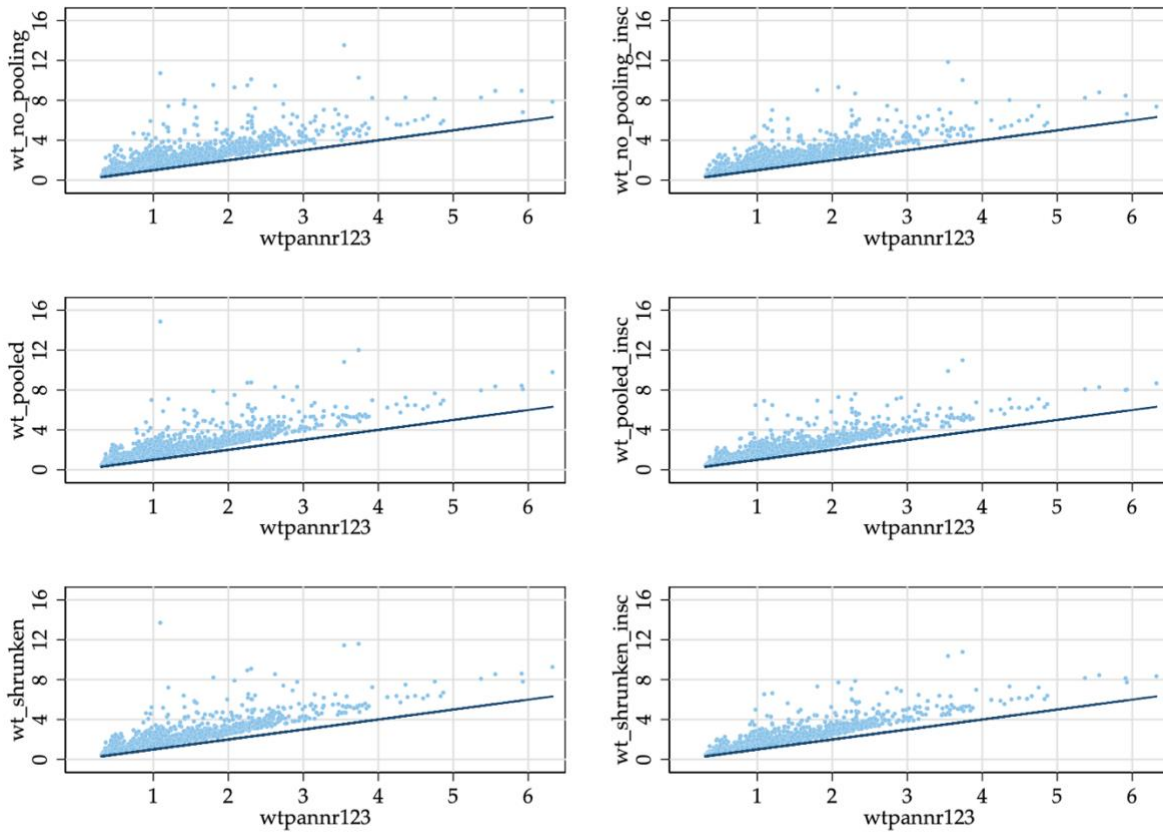
Figure 3. Scatterplots of year-specific attrition weights (top row), three-year-pooled attrition weights (middle row), and shrunken year-specific attrition weights (bottom row) by baseline NORC panel weight for the base-year eligible sample (left column) and the always in-scope, panel-eligible sample (right column)

Note: The dark line in each scatterplot is a 1-to-1 diagonal identity line for spatial reference only. It is less then 45 degrees because of the difference in scale for the horizontal and vertical axes.

Because the weights are right-skewed and based on an implicit multiplicative scale, it is common to transform weights by their natural logarithm in order to inspect their distributions with more care. Table 8 and Figure 4 are direct analogues to Table 7 and Figure 2, but for log-transformed weights, which spread out the weights on a linear scale.

Table 7. Descriptive statistics of log-transformed variants of the seven weights from Table 6

Name	Description	N	Mean	SD	Min	Max
lwtpannr123	NORC baseline panel weight	3,875	-0.150	0.531	-1.180	1.844
Based on the base-year eligible sample:						
lwt_no_pooling	Estimated separately by panel	3,875	0.269	0.573	-0.968	2.605
lwt_pooled	Estimated with panels pooled	3,875	0.274	0.563	-0.924	2.699
lwt_shrunken	Shrunk to pooled estimates	3,875	0.276	0.562	-0.932	2.618
Based on the sample that is in scope all waves:						
lwt_no_pooling_insc	Estimated separately by panel	3,875	0.212	0.576	-1.024	2.472
lwt_pooled_insc	Estimated with panels pooled	3,875	0.217	0.566	-0.971	2.397
lwt_shrunken_insc	Shrunk to pooled estimates	3,875	0.218	0.566	-0.980	2.377
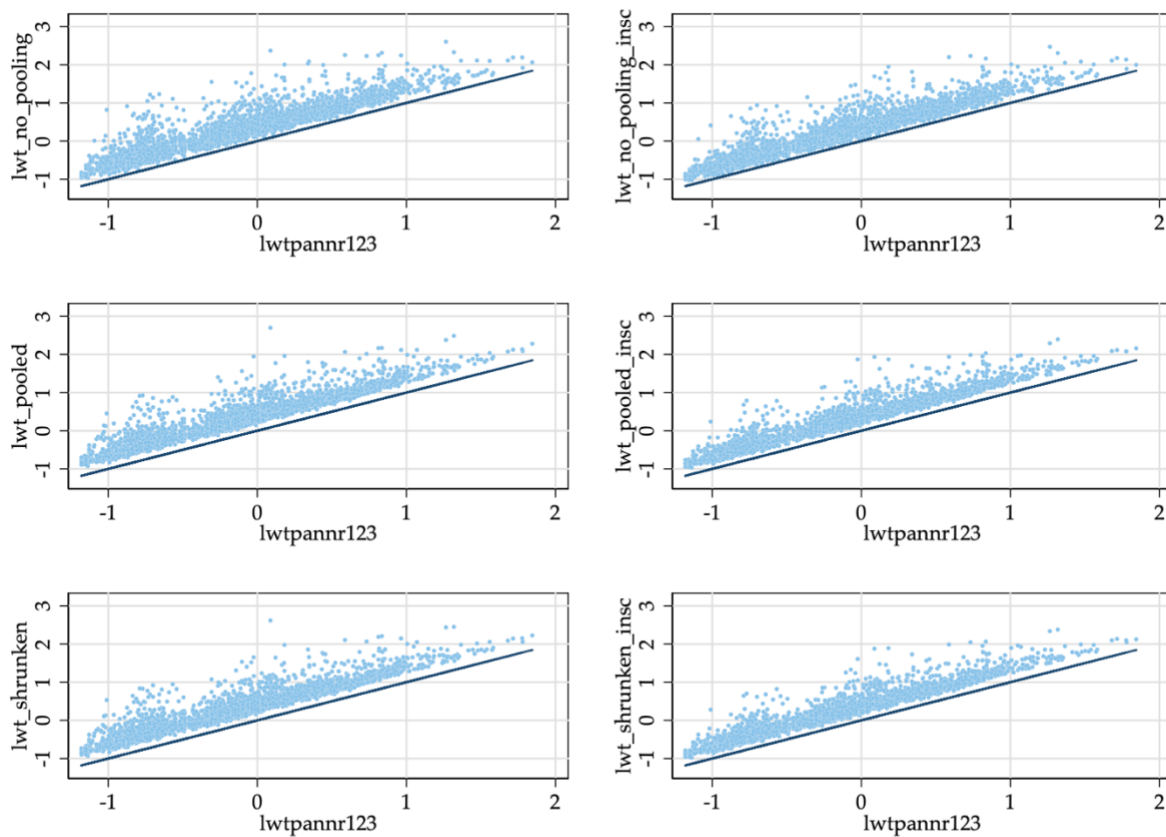
Figure 4. The same weights plotted for Figure 3 after log transformations

Note: The dark line in each scatterplot is a 1-to-1 diagonal identity line for spatial reference only.   It is less then 45 degrees because of the difference in scale for the horizontal and vertical axes.

For both Figures 3 and 4, the vertical dispersion of the weight values above the 1-to-1 diagonal identity line shows how the attrition adjustment weights have additional variance beyond the variance of NORC's baseline weight. If the estimated attrition probabilities had no variance, the attrition-adjustment ratios, $1/(1 - \hat{p}_i)$, would be equal to a single constant that would multiply the baseline weight uniformly. The weights would then fall on a new diagonal line above the diagonal line already plotted in each scatterplot (a steeper line for Figure 3 and a shifted but parallel line for Figure 4). Instead, the attrition-adjustment ratio, $1/(1 - \hat{p}_i)$, varies across individuals and generates vertical dispersion in each scatterplot.

How do the three rows of Figures 3 and 4 differ? The answer must be "not too much." One could imagine a pattern in which the three rows were very different, with the attrition weights based on panel-specific estimated probabilities having much more variance than those based on the pooled estimated probabilities (and with those based on the shrunken probabilities in between). That is what one might expect based on the patterns shown above in Figures 1 and 2. And this expected pattern is nonetheless present, just not as strongly as one might expect. The main reason that the weights do not differ as much as one might expect is that all six attrition weights are dominated by the common variation in NORC's weight, *wtpannr123*. In addition, this baseline weight has a very weak relationship with each of the six distributions of estimated probabilities of attrition. This can be seen in Figures 1 and 2, where the values of the probabilities presented in those figures are plotted with dots/circles that are proportional to each individual's value for *wtpannr123*. The sizes of the dots/circles appear randomly distributed across the scatterplots, or nearly so.

A careful observer will notice that the scatterplots differ by column, and this point is worth emphasis. When weights are constructed in an attempt, as for the left column, to support inference to the full base-year sample, the weights require more variance.

For a somewhat cheeky explanatory example, those respondents who were near death in the base year, but who nonetheless lived on for four more years in good enough health to sit for reinterviews, are asked by the underlying models to stand in for those who were near death in the base year but had worse luck (and/or health) and thereby succumbed to attrition by death. For the weights in the right column, there is less burden on these same near-dead panel respondents to represent others. The seemingly near-dead base-year respondents over the course of the panel only have to represent other near-dead base-year respondents who attrited for reasons other than death, not also near-dead base-year respondents who transitioned from near-dead to measured-dead. Of course, the patterns are more general, and death is simply one such dimension of note, and the most important one to consider. (A less cheeky example would consider the same pattern for those prone to emigration from the US over four-year intervals.)

More generally, the weights in the left column ask for more representation from panel respondents who are similar to those who become ineligible. In a sense, the weights in the left column must accomplish more, and they need to do so by relying on a smaller number of comparatively unusual respondents, such as the non-citizen, Spanish speaker in her nineties

who has been discussed at several points above. She had the largest estimated probability of attrition among non-attriters for the models that support inference to the full base-year sample. In Figure 3, her year-specific shrunken weight in the left column is the maximum value of 13.7 for the bottom left scatterplot. For the bottom right scatterplot, her weight falls to 6.5. For inference to the always-in-scope subsample, she does not need to stand in proportionally for those who died or left the US between 2008 and 2012.

GUIDANCE ON MERGING THE ATTRITION WEIGHTS

Downloading the entire set of code and data files that we have produced for this report will allow the reader to alter or refine our procedures for weight construction (see the repository, gss-treble-panel-weights, on GitHub; URL here). In addition, we provide an express route for those who simply wish to adopt one or more of our six weights for their own use (see the project page, "Treble Panel Weights for the GSS, 2006-2014," on the Open Science Framework; URL here).

We have produced csv and Stata data files that can be merged directly to the panel data released by NORC (and that can be downloaded at either of the two links above). The files are

Stata files:

gss-treble-panel-weights-wide.dta
gss-treble-panel-weights-long.dta

Comma-separated values files:

gss-treble-panel-weights-wide.csv
gss-treble-panel-weights-long.csv

In addition to the six weights, these files include variables for *year* and *id*. With these two variables, the weights can be attached to all individuals who have values for *wtpannr123*. A merge that uses these two variables as the key should result in a valid *n* for each of the merged weights that is exactly equal to the valid *n* for *wtpannr123*.

To minimize the odds of a merge error, we offer the following example Stata code (for version 15):

```
* GSS 2006-2010 panel

    use GSS_panel06w123_R6a.dta, clear
    rename *, lower
    clonevar year = year_1
    clonevar id = id_1
    sort year id
    merge m:m year id using gss-treble-panel-weights-wide.dta
    order year - _merge
```

```
        drop if _merge == 2
        codebook year - wtpannr123, c

* GSS 2008-2012 panel

        use GSS_panel08w123_R6.dta, clear
        rename *, lower
        clonevar year = year_1
        clonevar id = id_1
        sort year id
        merge m:m year id using gss-treble-panel-weights-wide.dta
        order year - _merge
        drop if _merge == 2
        codebook year - wtpannr123, c

* GSS 2010-2014 panel

        use GSS_panel2010w123_R6.dta, clear
        rename *, lower
        clonevar year = year_1
        clonevar id = id_1
        sort year id
        merge m:m year id using gss-treble-panel-weights-wide.dta
        order year - _merge
        drop if _merge == 2
```

The code above will need to be altered, for example, if new versions of the panel data are released by NORC. In addition, the "long" form of the merge files can be used to merge in weights to the three NORC panel data files if they are reshaped from "wide" to "long" format.

CONCLUDING DISCUSSION

In conclusion, we should first reassure any readers who may find the six weights we have constructed to be "too much of a good thing." Because attrition is not strongly patterned for the GSS treble panel, the alternative weights that we have developed are quite similar. All six weights have correlations with each other of 0.95 or higher. Their similarity is due, in large part, to their common component, NORC's baseline weight, *wtpannr123*, which accounts for most of their variance.

Our analysis also shows (because attrition is not strongly patterned) that simple analyses without adjustments for attrition may well, in the main, be sufficiently informative for many purposes. Where patterning is present, it is more likely to arise from the base-year interview experience, rather than demographic profiles. However, our discussion above shows how important it is to think carefully about the target of inference. We hope that it is clear that the weights that warrant inference to the full base-year panel sample should only be used with recognition of how doubly counterfactual they are. That target population is meaningful but artificial, since it is one for which those prone to death and emigration are prevented artificially from doing so. Inference to the always-in-scope target population does not require any such truly counterfactual attrition prevention. Yet, if one wants to align a panel analysis with a cross-sectional analysis of the 2006, 2008, or 2010 GSS – asking a question such as, "How would the cross-sectional sample evolve on outcome *XYZ* over four years *if* no one died and everyone

remained at residential addresses in the US?" – then the full base-year target population is the correct one to select. And, in that case, the weights based on the always-in-scope subsample would be inappropriate.

Altogether, if we had to choose one weight to use, we would first argue for the always-in-scope weights. Thus, it would be either *wt_no_pooling_insc*, *wt_pooled_insc*, or *wt_shrunken_insc*. In a sense, these weights achieve their value by shifting the analyst's perspective to value what the panel data are most suited for: modeling patterns among the subset of GSS respondents who remained eligible for the GSS treble panel over four years of study. Among these three weights, we would then (weakly) favor *wt_shrunken_insc*, which is very close to the pooled weight, but which gives the year-specific pattern of attrition just a bit of variance. Our second choice would then be *wt_pooled_insc*.

Alternative Perspectives and Possible Modifications

We conclude with some observations on how our weights might be altered by interested readers, even though our expectation is that none of these alterations would have a strong likelihood of improving on an analysis that uses the weights we have offered.

Alternative estimation methods. We use simple logit models to estimate the probabilities of attrition, and such models are a natural first choice. We also accept the individual-specific estimated probabilities from these models as worthy of using as they are. However, it is possible that the distributional assumption of the logit too strongly determines the particular estimated probabilities that individuals receive. Two (different) approaches could be taken. First, weighting class methods of the past would suggest using logit models but binning the estimated probabilities and forming corresponding groups of attrition "classes." Then, one would use each group's mean estimated probability for each individual. Second, one could seek to avoid the consequences of a fixed functional form by attempting to estimate non-parametric estimates of the probabilities, which would probably entail binning the values of the 21 predictors that we have specified for the logit models while eliminating others (such as those that appear mostly inconsequential: interviewer characteristics as well as the timing and dates of the base-year interviews). One could also combine these two approaches and bin the non-parametric estimates of the probabilities. Everything else about the weight construction would then remain the same.

Additional predictors of attrition. It is possible that modern data-mining methods could uncover an important determinant of attrition that we have not uncovered, and we confess to having relied heavily on the insights of Smith and Son (2010) and Frankel and Hillygus (2014). We are confident that we have not missed an obvious predictor of importance. Alas, the modern literature on data mining suggests rather convincingly that most investigators are overconfident about such matters.

Alternative specification control to avoid overfitting. As noted above, we prefer our shrunken weights because they protect against over-fitting that may afflict the specifications of

panel-specific models, using the simple rationale furnished by the empirical Bayes tradition of data analysis. We contemplated another strategy that is related but distinct. Accepting the pooling justification, one could think of the three specific panels as a cross-validation split. Using the same panel-specific models presented above, we would then calculate three estimated probabilities for each individual, using the three sets of coefficients from the panel-specific models. We could then take the individual-specific average of each set of three and anoint that value as the best estimate of each respondent's probability of attrition. For readers who know the cross-validation literature, this would not be optimal because, conditional upon accepting the pooling justification, it makes no sense to let baseline year define the cross-validation split. The more general alternative would therefore be to pool and embrace random cross-validation splits, using baseline year as an additional predictor, and then accept a single specification for all respondents. It is possible that these procedures would yield meaningfully different estimated probabilities, and if so, ones that would probably result in weights with less variance. Inferring the odds of change from the patterns shown in Figures 1 and 2, we do not think much change is likely. It is possible that this method, joined to a larger number of potential predictors (see our prior point), could lead to more substantial changes.

Extreme values and trimming. The weights we have constructed do not appear to be plagued by extreme values that have been produced by estimated probabilities far into the tails of their distributions. The most extreme value – for the Spanish speaking, non-citizen who participated in both reinterviews in her nineties – is quite sensible to us. And, by our reading of the literature on what should be regarded as extreme, her weight is not particularly large. In addition, when the realized sample (and thus implied target population) is narrowed to those who are always in scope, her weight falls into the range of the rest of the sample. Based on reasoning like this for the values of other respondents, we see no need to "trim" our weights to more conservative intervals by top-coding or bottom-coding. Nonetheless, investigators are welcome to trim as they see fit in order to assess the stability of their results, recoding the weights so that their maximum values are, for example, coded back to the 95th or 99th percentile, etc.

Using another method entirely. Because panel attrition can be considered a survey response pattern that generates unit-specific missing data for reinterview waves, an alternative perspective is to avoid using any type of attrition adjustment weight by instead (a) imputing item-specific values for attriters based on measures collected for non-attriters in reinterview waves and then (b) estimating models with multiple imputation methods. This strategy is pursued by Frankel and Hillygus (2014) for their consideration of the 2006 GSS panel, and it has considerable appeal. One might even argue that it is an ideal strategy to take. Even in this case, however, it is valuable to have defendable weights for preliminary analysis, and also for types of analysis where multiple imputation methods are not readily feasible.

References Cited

Frankel, Laura Lazarus and D. Sunshine Hillygus. 2014. "Looking Beyond Demographics: Panel Attrition in the ANES and GSS." *Political Analysis* 22:336-53.

Harter, Rachel, Stephanie Eckman, Ned English, and Colm O'Muircheartaigh. 2010. "Applied Sampling for Large-Scale Multistate Area Probability Designs." Pp. 169-97 in *Handbook of Survey Research*, edited by P. V. Marsden and J. D. Wright. Bingley, UK: Emerald Group.

Hout, Michael and Orestes P. Hastings. 2016. "Reliability of the Core Items in the General Social Survey: Estimates from the Three-Wave Panels, 2006-2014." *Sociological Science* 3:971-1002. doi: 10.15195/v3.a43.

Kish, Leslie. 1965. *Survey Sampling*. New York: Wiley.

Levy, Paul S. and Stanley Lemeshow. 2008. *Sampling of Populations: Methods and Applications*, 4th Edition. Hoboken: Wiley.

Smith, Tom W., Michael Davern, Jeremy Freese, and Stephen L. Morgan. 2019. *General Social Surveys, 1972-2018, Cumulative Codebook*. Chicago: NORC.

Smith, Tom W. and Ben Schapiro. 2017. *General Social Surveys, 2006-2014 Panel Codebook*. Chicago: NORC.

Smith, Tom W. and Jaesok Son. 2010. "An Analysis of Panel Attrition and Panel Change on the 2006-2008 General Social Survey Panel." GSS Methodological Report No. 118, National Opinion Research Center, Chicago, Illinois.

Thompson, Steven K. 2012. *Sampling*, 3rd Edition. New York: Wiley.

Valliant, Richard, Jill A. Dever, and Frauke Kreuter. 2013. *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.