O'Muircheartaigh, C., English, N., Pedlow, S., & Kwok, P.K. (2014). Sample design, sample augmentation, and estimation for Wave 2 of the national social life, health, and aging project. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, *69(8)*, S15–S26, doi:10.1093/geronb/gbu053

Sample Design, Sample Augmentation, and Estimation for Wave 2 of the NSHAP

Colm O'Muircheartaigh,[1] Ned English,[2] Steven Pedlow,[2] and Peter K. Kwok[2]

[1]University of Chicago Harris School and NORC and
[2]NORC at the University of Chicago, Illinois.

Objectives. The sample for the second wave (2010) of National Social Life, Health, and Aging Project (NSHAP) was designed to increase the scientific value of the Wave 1 (2005) data set by revisiting sample members 5 years after their initial interviews and augmenting this sample where possible.

Method. There were 2 important innovations. First, the scope of the study was expanded by collecting data from coresident spouses or romantic partners. Second, to maximize the representativeness of the Wave 2 data, nonrespondents from Wave 1 were again approached for interview in the Wave 2 sample.

Results. The overall unconditional response rate for the Wave 2 panel was 74%; the conditional response rate of Wave 1 respondents was 89%; the conditional response rate of partners was 84%; and the conversion rate for Wave 1 nonrespondents was 26%.

Discussion. The inclusion of coresident partners enhanced the study by allowing the examination of how intimate, household relationships are related to health trajectories and by augmenting the size of the NSHAP sample size for this and future waves. The uncommon strategy of returning to Wave 1 nonrespondents reduced potential bias by ensuring that to the extent possible the whole of the original sample forms the basis for the field effort. NSHAP Wave 2 achieved its field objectives of consolidating the panel, recruiting their resident spouses or romantic partners, and converting a significant proportion of Wave 1 nonrespondents.

Key Words: Longitudinal survey design—NSHAP Wave 2—Panel attrition—Panel survey—Panel weighting—Partner reports.

NATIONAL Social Life, Health, and Aging Project (NSHAP) was designed ab initio as a cohort study; the intention was to follow a national probability sample of older Americans (aged 57–85) from recruitment in 2005 through to their final life stages. Thus, the design and estimation issues that arose in Wave 2 (conducted in 2010) primarily dealt with linkage of data across the two waves and with designing weights to take into account attrition between waves. In addition to these typical concerns, however, Wave 2 of NSHAP introduced two important innovations that have important implications for Wave 2 and cross-wave analyses. First, the scope of the study was extended to include the coresident romantic partners of those selected in Wave 1, for which only one respondent was selected per household; the Wave 1 sample member was designated the prime respondent. Including these partners changes the probabilities of selection of all Wave 2 respondents, as eligible individuals now have two different routes into the sample—as primes and as the partners of primes; a randomized methodological investigation built into Wave 2 provides valuable information but further complicates the analysis. Second, we revisited the selected sample members who failed to respond in Wave 1. As our estimates for Wave 1 incorporated a weighting correction for differential nonresponse, the successful recruitment of a proportion of these nonrespondents has implications for the Wave 2 weights.

First, the Wave 1 sample design is described briefly. The elements of the research design for Wave 2 with implications for data analysis are then presented. The rationale for the enhanced sample design is described, and the response rates for the different components of the design are presented. The results of the field implementation of the partner recruitment, and the embedded experiment, are given. In the final sections, the weighting system is described, and some exemplars of the standard errors of the sample estimates and their design effects are given. The most significant results are summarized in the Conclusion section.

Wave 1 Sample Design

The design for the Wave 1 NSHAP sample was an area-based national probability sample of older household-resident Americans. A full description of the design can be found in O'Muircheartaigh, Eckman, and Smith (2009); a brief summary is given below. Geographic areas were selected with probabilities proportional to their sizes—typically in two stages; subsequently, a sample of households was selected in each of these areas. The three stages together determine the probabilities of selection of the individuals in

Received November 4, 2013; Accepted April 11, 2014
Decision Editor: Mark Hayward, PhD

the study. This design is a classic multistage area probability sample (for details on this class of sample designs, see Harter, Eckman, English, & O'Muircheartaigh, 2010). The probabilities of selection for the households were designed to provide approximately equal probabilities of selection for all eligible individuals. The selection of households and individuals for NSHAP was carried out following screening of all households in the selected areas.

The original NSHAP target population was adults aged 55–85 years in 2004. As individuals in this age range can be found in only 30% of U.S. households, identifying such households through screening would involve an expensive and time-consuming field operation. However, as NSHAP was being planned, the Health and Retirement Study (HRS, also funded by the National Institute on Aging) was about to embark on the recruitment of a new cohort. Through an innovative collaboration between NSHAP and HRS (and between the National Opinion Research Center [NORC] and the Institute for Social Research, the respective survey organizations), the screening for both surveys was carried out as a single operation, with substantial cost saving. As HRS interviewers screened households in the selected segments for individuals eligible for their survey, from February to November 2004, they also identified individuals who were eligible for NSHAP. This sharing of field resources allowed NSHAP to have a much larger sample size than would otherwise have been possible. The collaboration did require, however, that NSHAP redefines its target population to adults aged 57–85 years. Consequently, the HRS and NSHAP populations were very nearly nonoverlapping; the minor overlap is described in O'Muircheartaigh and colleagues (2009).

Thus, the NSHAP sample is built on the foundation of the national household screening carried out by the HRS for the recruitment of their 2004 cohort (adults born between 1948 and 1953) and 2010 cohort (adults born between 1954 and 1959). In 2004, HRS compiled a complete listing of housing units selected under a multistage area probability sample design, with oversampling of areas with relatively high density of African Americans and Latinos. Households identified as having at least one member born prior to 1948 were made available to NSHAP. (In cases where there was an HRS-eligible individual living in a household with an NSHAP-eligible individual who was not his/her partner, those households were not made available to NSHAP.) The study design aimed to equalize the projected number of respondents by gender for three age groups (57–65, 66–74, 75–84) and retained minority oversampling built into the HRS sample.

The resulting frame comprised 5,922 households containing 7,769 age-eligible individuals. An iterative selection procedure was developed, which minimized variability in selection probabilities (and hence weights), given the objective of equalizing the number of selected individuals in each of the six age by gender cells. Our approach maximized the precision of the survey estimates for a given sample size. We selected a final sample of 4,400 persons (one from each of 4,400 households) from this frame. We estimated that this sample would generate approximately 3,000 completed interviews; at the end of the fieldwork, the number of completed cases was 3,005. Wave 1 achieved a final overall response rate of 75.5%, with response rates exceeding 70% for each age–gender category and for each ethnicity. Approximately 6% of all interviews were completed in Spanish. Table 1 presents final case dispositions. Data are publicly available (NSHAP Wave 1: Waite, Linda J., Edward O. Laumann, Wendy Levinson, Stacy Tessler Lindau, and Colm A. O'Muircheartaigh. National Social Life, Health, and Aging Project (NSHAP): Wave 1. ICPSR20541-v6. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2014-04-30. doi:10.3886/ICPSR20541.v6.).

Wave 2 Research Design

Our first aim in NSHAP Wave 2 was to increase the scientific value of our Wave 1 data set by revisiting respondents 5 years after their initial interviews to examine changes in key measures of health, sexuality, and social relationships. In addition, the scope of the study was expanded by collecting data from coresident spouses or romantic partners, something that was not attempted at Wave 1. We describe the Wave 2 sample design below.

Sample Design and Enhancement

The Wave 2 data collection focused on the dynamic relationships between social life, sexuality, and health in the lives of older adults. The addition of partner interviews permitted in-depth exploration of the joint production of health for coresident spouses or romantic partners. The Wave 2 sample thus included three types of respondents: (a) Respondents who

Table 1. NSHAP Final Sample Dispositions for Wave 1

		Total sample		Eligible cases	
Disposition category	Cases (N = 4,400)	Unweighted	Weighted	Unweighted	Weighted
Interviewed	3,005	68.3	69.2	74.8	75.5
Final refusal	990	22.5	22.0	24.6	24.0
No contact	9	0.2	0.1	0.2	0.2
Moved	3	0.1	0.2	0.1	0.1
Other	10	0.2	0.2	0.2	0.2
Out-of-scope	383	8.7	8.3	—	—

Note. NSHAP = National Social Life, Health, and Aging Project.

participated in the first wave were reinterviewed. These individuals constitute the panel component of NSHAP and allow population-based longitudinal analysis of health, sexuality, and social relationships. (b) Eligible individuals selected for Wave 1 who did not respond were approached again in Wave 2. These two groups together comprise the *prime* respondents. (c) The partners of the prime respondents were also included.

A common challenge in longitudinal studies is that the gradual attrition of the panel across waves leads quickly to inadequate sample size and potential bias. In NSHAP, we aimed to minimize the attrition effect by approaching individuals who chose not to participate—or were not available to participate—in Wave 1. NORC had already achieved some success with this approach in other surveys: (a) In the *Resident Relocation Survey* for the MacArthur Foundation, a panel survey of Chicago Housing Authority leaseholders, NORC achieved a response rate in Wave 2, a year later, of more than 50% for Wave 1 nonrespondents. (b) In a panel survey of inner-city residents in the Annie E. Casey Foundation's *Making Connections* project, among Wave 1 nonrespondents, NORC achieved Wave 2 response rates of between 40% and 60%, again a year later.

We wished to ameliorate the problem of permanent first wave nonresponse in NSHAP, even if we could not avoid it completely, by returning in Wave 2 to Wave 1 nonrespondents. The NSHAP sample presents somewhat different challenges from our previous NORC experience, and our expectation was accordingly lower. In particular, intensive efforts had been made in Wave 1 to obtain cooperation, the gap between the waves was to be 5 years, and the respondents were transitioning to an age group with lower expected response rates; thus, the prognosis for success was somewhat less positive. However, there is substantial evidence that a good deal of nonresponse is situational and transient rather than immutable. The standard fieldwork protocol involves repeated visits to reluctant sample members; such repeat visits frequently yield completed interviews, and the strategy of returning to nonrespondents from an earlier wave of a cohort is based on an extrapolation of this process. In earlier studies at NORC (survey of Chicago Housing Authority leaseholders; Making Connections [unpublished field reports]), we have obtained response rates between 30% and 50% for such cases. Given the special considerations described earlier, we projected a 20% response rate for this group. We are aware of the cost-benefit tradeoffs implicit in this approach: the benefit is the enhancement of the representativeness of the NSHAP Wave 2 sample; the direct cost arises from the interviewer training and time necessary to obtain the interviews, the indirect cost is in the potential aggravation of the interviewers by dealing with sample members who are generally less agreeable than the norm.

A total of 1,012 cases were coded as noninterviewed respondents in Wave 1, which covers all cases believed to be eligible and who did not cooperate. Those sampled for Wave 1 but who were deemed hostile when contacted were removed from the Wave 2 sample. The remaining 824 nonhostile Wave 1 nonrespondents were approached for participation in NSHAP in Wave 2.

Besides returning to Wave 1 nonrespondents, another major innovation at Wave 2 was the plan to interview cohabitating spouses or romantic partners. For the Wave 2 partner study design, new Wave 2 respondents and those who are reinterviewed following their participation in Wave 1 together constitute the prime respondents; these are members of the original NSHAP sample. Because our hypotheses focus on the importance of the intimate partner relationship, our initial design planned for each prime respondent to seek to interview the current *coresident spouse or romantic partner*. (Following discussion with our colleagues in the field division, we had decided not to attempt interviews with the partners of Wave 1 nonrespondents. During the fieldwork, having had a positive response from some of those nonrespondents, we did approach a subset of their partners.) For Wave 1, we decided to include only one respondent per household. This is the standard approach for many household surveys for three reasons. First, there is considerable contamination potential as described below. Second, where the interest is in estimates across the whole sample, including additional household members typically leads to a substantial increase in the design effect, and consequently the standard errors for a given sample size. Third, conditional response rates for second and subsequent household members are not always high. NSHAP differs from most household surveys in that the almost all analyses are gender specific, most partners are of different gender than the primes, and thus the within-gender design effects are unaffected by their inclusion. The contamination risk remained a concern however.

We were concerned that the introduction of partners might influence our results in two ways: (a) the response rate for the prime respondents might be decreased and (b) the responses of the primes themselves to sensitive questions might be affected. Consequently, an evaluation experiment (described below) was incorporated into the fieldwork.

The first advantage of the partner study design is that it allows us to examine if and how romantic or sexual relationships promote better health trajectories. Analyses that require such data include: (a) paired analyses, involving within-couple comparisons to examine gender differences or control for unobserved household-level characteristics; (b) the estimation of the within-couple correlation in various health outcomes, controlling for individual-level factors; and (c) analyses in which the couple is the unit of analysis, using parallel data from both partners and/or derived variables capturing features of the couple (measures of similarity or difference between partners are examples). Projecting forward to Wave 3, returning to prime respondents and their spouses or romantic partners in a possible third wave of NSHAP would also allow longitudinal analysis of partnership characteristics and relationship quality alongside health trajectories.

A second (and independent) advantage to the inclusion of partners is the effect on sample size. Projections based on data from Wave 1 suggested that reinterviewed respondents would yield estimated 1,827 coresident

romantic partners. *Age-eligible partners* (ages 57–85 at the time of Wave 1) constitute additional respondents for cross-sectional analyses and increase the power of the research design. Figure 1 summarizes the age distributions of coresident partners, as indicated in Wave 1 interviews with referent respondents. Potentially, the sample size could be augmented by more than 1,500 age-eligible individuals. With a response rate of 75%, this would provide an additional 1,100 cases for cross-sectional analysis in Wave 2. The impact on the precision of estimates, particularly for gender-specific estimates, would be considerable. For gender-specific estimates, the sample of women (and the sample of men) would be some 40% larger; as most partners are of the opposite gender, these additional respondents would represent additional households, and the intra-household correlation coefficient would be effectively zero. Consequently, the effective sample size would be increased proportionately, reducing sampling variances by 30% (and standard errors by 15%). For comparisons between men and women (where both partners would be included in the analyses), the within-household correlation would in fact reduce the variance of the estimates below that for a comparison of samples of the same size of men and women from different households.

Furthermore, augmentation of the sample would provide a larger base for a future Wave 3 of NSHAP. At the time of Wave 2 fieldwork, we anticipated that about 12% of male respondents and 9% of female respondents from Wave 1 would have died (see Table 2). The sample size will also have been depleted due to residential mobility and transitions to care facilities. If we were to return only to the Wave 1 respondents, the expected number of completed interviews in Wave 2 would be 2,173.

One of the operational challenges in adding partners to NSHAP for Wave 2 was creating a uniform method for identifying spouses and romantic partners as respondents (see Jaszczak et al., 2014). Prime respondents identified partners through a series of questions administered during the Social Network Roster section of their own interview. Once an eligible cohabiting spouse or romantic partner was identified, appropriate materials for later administration to the partner were generated by the system, and the partner was notified of our intention to approach the partner (see also, below, the discussion of the partner experiment). (The wording of the interviewer's statement was: "We know that our partners have an important impact on our lives. To better understand the roles our partners play in shaping our health and social relationships, we may want to ask your partner to participate in an interview. Just as I asked you about your own attitudes and behaviors, I would like also to ask your partner about their own attitudes and behaviors. If your partner decides to participate, please know that I will not share your confidential survey answers with your partner, or vice versa. All of the answers will remain confidential.[PAUSE] Do you have any questions before I approach your partner to participate?")

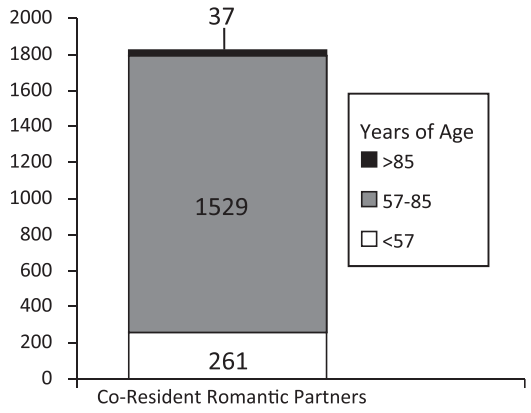

Figure 1. Age distribution of Wave 1 partners.

After adjusting for mortality and morbidity projections and assuming that 5% of respondents would be out-of-scope (primarily due to incapacity), the augmentation of the Wave 2 sample with coresident romantic partners as well as Wave 1 nonrespondents and their coresident partners was expected to add a total of about 1,214 eligible respondents (see Table 2). After applying projected response rates, we anticipated a total of 3,387 completed cases in NSHAP Wave 2.

Both mortality and incapacity are important outcomes, and the data collection strategy was modified at Wave 2 to incorporate important information on respondents who drop out due to disability or death. In the case of incapacity, information was collected from proxy respondents where possible. For nonrespondents (noncontacts), we attempted to establish mortality by interviewing proxies. Data are publicly available (NSHAP Wave 2: Waite, Linda J., Kathleen Cagney, William Dale, Elbert Huang, Edward O. Laumann, Martha K. McClintock, Colm A. O'Muircheartaigh, L. Phillip Schumm, and Benjamin Cornwell. National Social Life, Health, and Aging Project (NSHAP): Wave 2 and Partner Data Collection. ICPSR34921-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2014-04-29. doi:10.3886/ICPSR34921.v1.).

Response Rates

Of the 4,017 potential respondents selected for Wave 1 of NSHAP, 3,325 were still eligible in 2010 at the time of the Wave 2 fieldwork. Of these, 2,552 were successfully interviewed, an unweighted response rate of 73%; the appropriately weighted response rate (taking into account the original probabilities of selection) was 74%.

There is no single response rate that represents adequately the outcome of the second wave of a longitudinal study; below, we discuss two types of response rates: (a) the *conditional* response rates for different subsets of the sample and (b) the *unconditional* response rates for different scenarios across the two waves. We discuss in particular the impact on the unconditional response rate of our decision to return to Wave 1 nonrespondents.

Table 2. Estimated Number of Completed Cases, NSHAP Wave 2

	Number in sample		Number surviving at start of wave		Number of eligible respondents[a]		Expected response rate		Estimated number of completed cases	
Wave 2 group	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀
Wave 1 completes	1,455	1,550	1,282	1,409	1,218	1,339	85%	85%	**1,035**	**1,138**
Coresident romantic partners	741	1,086	576	904	547	859	70%	70%	**383**	**601**
Wave 1 nonrespondents (NIRs)[b]	463	523	405	468	385	445	20%	20%	**77**	**89**
Coresident romantic partners of Wave 1 nonrespondents[c]	—	—	—	—	35	57	70%	70%	**24**	**40**
Total[d]	**2,659**	**3,159**	**2,263**	**2,781**	**2,185**	**2,700**	—	—	**1,519**	**1,868**

Notes. NIR = Noninterviewed Respondents; NSHAP = National Social Life, Health, and Aging Project.
[a]Eligible respondents are estimated by assuming that 5% of the number of individuals surviving at the start of Wave 2 will be out-of-scope.
[b]Individuals whose gender was not reported were excluded from the calculations and reduced the NIR sample from 1,012 to 986. Projections use these 986 individuals.
[c]The number of coresident romantic and nonromantic partners is based on the 166 projected completed cases from the Wave 1 NIRs.
[d]The total number of cases (actual or estimated) in each category.

Table 3. Conditional Response Rates (RRs) for NSHAP Primes (Members of the Original Wave 1 Sample) at Wave 2

	Wave 1 respondents		Wave 1 nonhostile nonrespondents		Wave 1 hostile respondents	
Category	Cases	RR	Cases	RR	Cases	RR
Born 1920–1929	823	82%	274	20%	61	0%
Born 1930–1938	985	91%	274	25%	60	0%
Born 1939–1947	1,197	91%	276	30%	67	0%
All	3,005	89%	824	26%	188	0%

Note. NSHAP = National Social Life, Health, and Aging Project.

The conditional response rate for Wave 1 respondents at Wave 2 is the response rate at Wave 2 among those who responded to Wave 1. This rate represents the success of the survey organization in retaining members of the panel. Overall, the conditional response rate was 89%, four percentage points higher than our target (predicted) rate (see Table 3). Given the complex, time-consuming, and intense nature of the in-home experience and the aging of the panel, this represents a considerable achievement. The increasing difficulty of obtaining a response as the panel members age can be seen in the rates for the three age groups: 91% for those born in 1939–1947 and 1930–1938, but only 82% for those born 1929 or earlier.

Schoeni, Stafford, McGonagle, and Andreski (2013, pp. 64–65) present a summary of reinterview response rates for six major national panel surveys, which provides context for these results. In general, the year-on-year conditional response rates for these surveys are above 90%, and often above 95%; generally, the lowest rates are in the early years of the panel. Were we to extrapolate from an annual 90% rate, we would predict a 5-year rate of 59%; with a 95% rate, we would predict a 5-year rate of 77%; with a 98% annual rate, we would project a 5-year rate of 90%. The NSHAP 5-year rate of 89.4% is at the very upper end of these possibilities and is particularly noteworthy given (a) the quite elderly population represented by the sample and (b) the fact that this 5-year period incorporates the early stages of the panel's existence.

The decision to return to Wave 1 nonrespondents was based partly on evidence of success from other surveys and partly on the concern that writing off Wave 1 nonrespondents (a quarter of the sample) would inevitably lead to a more rapid attrition in the panel over time. There were 1,012 nonrespondents in Wave 1, defined as those eligible for inclusion in the survey who chose not to participate. Of these, 188 were classified as "hostile" nonrespondents; these were excluded from the Wave 2 fieldwork. The remaining 824 nonrespondents were approached; the interviewers were aware that they had been nonrespondents in Wave 1.

Among those still eligible within the 824 Wave 1 nonhostile nonrespondents, 26% responded in Wave 2. This was six percentage points higher than the predicted (target) response rate. As a result, the Wave 2 responding sample was increased by 161 individuals. Not only do they augment the sample on which Wave 2 cross-sectional estimates can be based but they also augment the panel as it moves forward to Wave 3; they do not, of course, have measures for Wave 1.

The same pattern across age groups is evident in Table 3 for the conditional response rates for Wave 1 nonrespondents. Our success in converting Wave 1 nonrespondents has a monotonic relation to age; for the youngest cohort, we were successful in converting 30%; for the middle cohort, 25%; for the oldest cohort, we obtained cooperation in Wave 2 from only 20% of the nonhostile nonrespondents from Wave 1.

The *unconditional response rate* for Wave 2 is the proportion of the target population (eligible adults born between

1920 and 1947) that is represented by the responding sample in Wave 2. This response rate can be computed by combining the responses across the Wave 1 respondents and the Wave 1 nonrespondents. For Wave 2, this rate is 73.9%; this is remarkably close to the Wave 1 response rate of 75.5%, indicating very slight attrition between Wave 1 and Wave 2.

It is worthwhile considering the composition of this response rate. Had we gone back only to those who had responded in Wave 1, the final unconditional response rate would have been 69%. The Wave 1 response rate was 75.5%. Attrition (loss of respondents due to nonresponse) between Wave 1 and Wave 2 was 6.25% of the total eligible sample. The decision to return to nonhostile nonrespondents generated an unconditional 4.6% of the eligible sample (two thirds of the attrition). Thus, the net attrition in Wave 2 represents less than 2% of the eligible population.

Table 4 shows the overall response rate and the response rate for each of the three age cohorts; in each case, the composition of the response rate is presented: (a) the response rate for those who responded at both waves and (b) the augmentation obtained by converting Wave 1 nonrespondents. The pattern across age groups is familiar. Overall, the response rate is highest for the youngest cohort and lowest for the oldest cohort. The augmentation due to conversion of nonrespondents is very similar across cohorts. We are not able to determine to what extent this is an age effect and to what extent it is a cohort effect. If we were to extrapolate both these results from Wave 2 through Wave 3, the impact is enhanced. Reasonable assumptions in about conditional response rates for Wave 3 might be:

For those who responded to both Wave 1 and Wave 2: 95% conditional response rate;
For those who responded to Wave 2 only: 89% conditional response rate;
For those who responded to Wave 1 but refused at Wave 2: 30% response rate.

Using these rates, were we not to return to nonrespondents in either wave, our unconditional response rate for Wave 3 would be:

$$\text{Wave 3 RR} = 69.5\% \times 0.95 = 66\%$$

With our strategy of returning to nonrespondents after one wave of nonresponse, the unconditional response rate for Wave 3 would be:

$$\text{Wave 3 RR} = 0.95 \times 69.5\% + 0.89 \times 4.64\% + 0.30 \times 6.5\% = 72\%$$

Once again the rate of attrition is more than halved.

The Partner Study Evaluation

Given the sensitive nature of some of the questions regarding health behaviors and relationship quality, the NSHAP team was concerned that introducing partner interviews could discourage response from the primes or potentially introduce a bias to their responses. These concerns had been influential in the original decision not to include partners in the design of NSHAP Wave 1. The NSHAP Advisory Board supported the recommendation that an evaluation be conducted of the impact of introducing partner interviews. Consequently, a randomized experiment was designed to test a number of basic questions. Though there is no coherent theory in this area, we formulate also tentative hypotheses with respect to these questions based on the conventional wisdom.

For primes:

1. Whether the introduction of partner interviews affects the response rates of the Wave 1 respondents (primes), both to the in-person interview and for the Leave-Behind Questionnaire (LBQ). Hypothesis 1: the introduction of partner interviews will reduce the response rate of the primes
   1a. A subsidiary question is whether the timing of the request to interview the partner affects the response rates. The direction of the effect is not clear a priori.

For partners:

2. Whether the timing of the request affects whether we find partners and whether, when found, they respond. Hypothesis 2: the timing of the request will have an effect; the direction of the effect is not clear a priori.

Effect on responses:

3. Whether the introduction of partner interviews or the timing of the request affects the responses of those primes who respond to Wave 2, either to the in-person interview or the LBQ. We plan to address this issue in a later paper.
   3a. We do not address the contamination of partner reports; indeed only the differential effect by timing can

Table 4. Unconditional Response Rates (RRs) for NSHAP Primes (Members of the Original Wave 1 Sample) at Wave 2

Category	Response rate due to those responding to both waves	Response rate augmentation due to converted nonrespondents	Overall unconditional response rate
Born 1920–1929	60.4%	4.3%	64.7%
Born 1930–1938	68.9%	4.7%	73.6%
Born 1939–1947	72.5%	4.7%	77.2%
All	69.3%	4.6%	73.9%

Note. NSHAP = National Social Life, Health, and Aging Project.

be addressed in this study, and we will explore this issue in a later paper.

For general inferences:

4. A final question arises as to how the responding partners can properly be incorporated in the individual (as distinct from partnership) estimates from NSHAP.

Response Rates Among Primes

Wave 1 respondents were first categorized according to whether they had a partner at the time of the Wave 1 interview. Those who had a partner at Wave 1 were assigned to one of three experimental groups: (a) Control group—partners not selected; (b) Tell Before group—partner selected and prime informed at the beginning of the Wave 2 interview; (In the "Tell Before" condition, the prime was informed that we would wish to interview her/his partner as soon as the interviewer had confirmed that the prime had (or still had) a partner. This usually occurred during or at the end of the *Social Network Roster* section of the interview, quite early in the visit.) (c) Tell After group—partner selected but prime not informed until the end of the Wave 2 interview. The design is illustrated in Figure 2; the allocation of sample sizes is given in Tables 5 and 6.

Tables 5 and 6 compare the response rates of the primes across the different experimental groups. As the assignment rules differed according to whether the prime had reported a partner in Wave 1, the comparison is made separately for different Wave 1 partner status. A difference between the response rates for the "Not Selected" cases and the "Partner" cases would indicate that knowing the partner would be approached could color the interviewer's expectations (and therefore behavior) in a way that influenced their effectiveness in inducing cooperation from the prime. A difference between the "Tell Before" and "Tell After" cases would indicate that the interviewer could have a different level of concern (approaching the case) about the stage in the prime interview that the topic of interviewing the partner would be broached with the respondent. In presenting these comparisons, we use unweighted response rates for simplicity; the weighted response rates are almost identical.

Each of the 1,835 respondents who had reported a partner in Wave 1 was allocated to one of three conditions: "Tell Before," "Tell After," and a control group. Table 5 shows the response rates for the groups; these are the response rates for the primes themselves. There is no effect either of being invited to include one's partner or of the timing of the invitation. The response rates are all essentially identical. This allays a major concern, encapsulated in Hypothesis 1, that the decision to recruit partners in Wave 2 would depress the response rates of the primes. We also reject Hypothesis 1a for this set of respondents, that the timing of the request has an effect.

Each of the 1,170 respondents who had reported no partner in Wave 1 was allocated to one of only two conditions: "Tell Before" and "Tell After." Table 6 shows the results. This provides a separate test of Hypothesis 1a for this distinct set of primes. There was a statistically significant difference in the response rates; interestingly, the response rate was higher for the "Tell Before" group. There are two competing forces that might be at play here: on the one hand, earlier disclosure of our intention might reduce the willingness of the prime to be interviewed, but on the other hand, the effect on the interviewer of knowing that our intention would be revealed later might diminish the interviewer's job performance. These cases (all those without partners in Wave 1) generated very few partners, only 42 in all. The small number of partners involved makes the difference substantively insignificant in terms of partner analysis.

Response Rates Among Partners

Table 7 presents response rates for partners who were identified (or confirmed) during the Wave 2 interview. The

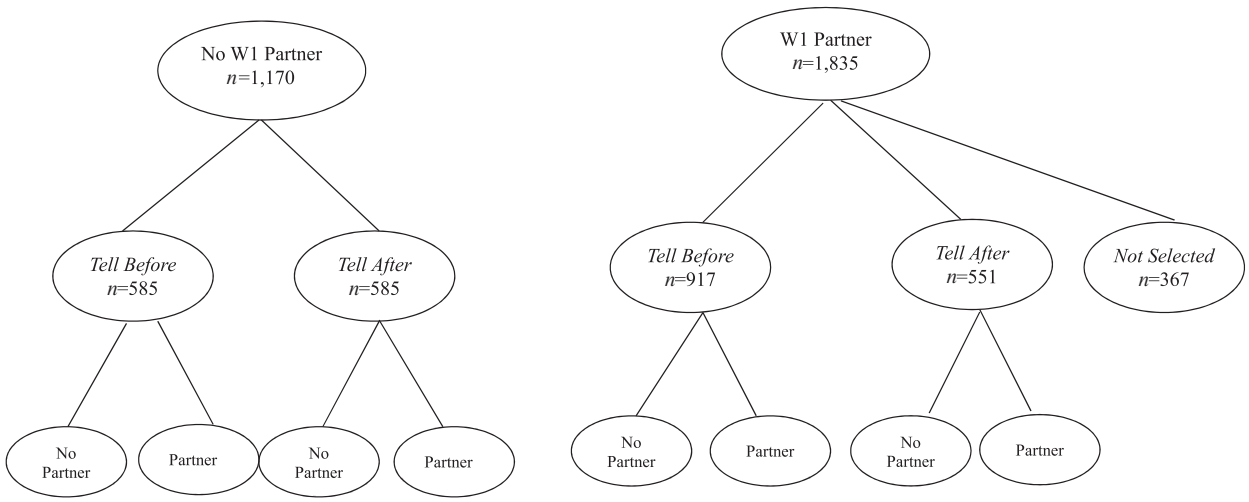

Figure 2. Primes (n = 3,005).

Table 5. Prime Response Rates With Wave 1 Partner

Wave 1 partner	Not selected	Tell before	Tell after	Total
Wave 2 respondents with no Wave 2 partner	44	98	53	195
Wave 2 respondents with Wave 2 partner	256	624	392	1,272
Wave 2 nonrespondents	33	79	47	159
Wave 2 out-of-scope	34	116	59	209
Total	367	917	551	1,835
Response rate	90.1%	90.1%	90.4%	90.2%

Table 6. Prime Response Rates With No Wave 1 Partner

No Wave 1 partner	Tell before	Tell after	Total
Wave 2 respondents with no Wave 2 partner	380	371	751
Wave 2 respondents with Wave 2 partner	32	11	43
Wave 2 nonrespondents	55	77	132
Wave 2 out-of-scope	118	126	244
Total	585	585	1,770
Response rate	88.2%	83.2%	85.7%

overall (conditional) response rate was 84.3%, substantially higher than the 70% we had assumed in our design. The response rates were essentially identical for men and women. There was some variation by age, though not as pronounced as the variation for primes; again, the oldest respondents had the lowest response rate, though only three percentage points below the mean. There were some partners who were outside the eligible age range, mostly younger women and older men; their response rates were slightly lower than those for the age-eligible partners. The biggest difference was for partners of Wave 1 nonrespondents. Only two thirds of these responded. This may be explained by the fact that the field effort for those partners was considerably shorter than that for the others, due to (a) the late decision to pursue them and (b) the fact that Wave 1 nonrespondents (the primes corresponding to those partners) were in general interviewed later in the field period. Excluding these exceptional cases from the calculations, the response rates are about a point and a half higher. Table 7 gives the detailed results.

Interestingly, there was a statistically significant difference of 3% (in favor of the "Tell Before" group) between the response rates for the partners in the "Tell Before" and "Tell After" conditions. This parallels the weak indication of a difference in favor of the "Tell Before" condition among primes who did not have a partner at Wave 1. We are investigating this further in a separate study of the partner experiment.

Weighting

Construction of Weights

The complex design of NSHAP requires that the data be weighted in order to provide unbiased estimates of population parameters. In this section, we use the names given to the weighting variables in the public use data set, so that analysts can easily connect this narrative to their analyses.

The NSHAP public use data set contains one weight (weight_adj) that can be used for all analyses of all Wave 2 respondents. As described earlier, the Wave 2 respondents are either prime respondents (those who were selected into the Wave 1 sample and responded in Wave 2) or partners (the romantic partners sharing a household with a prime respondent). Three additional weights were created for special uses: weight_sel (this weight adjusts for selection probabilities, but not for nonresponse), weight_ageelig (this weight is zero for partners who were age ineligible for Wave 1), and weight_couple (for analyses of the 957 couples who both completed a Wave 2 interview).

The steps involved in calculating the weights for the Wave 2 respondents build on the procedure for Wave 1 (as described in O'Muircheartaigh et al., 2009). There is an important difference, however. In Wave 1, only one respondent per household was selected; in Wave 2, some households provide two respondents – the prime and the partner. Therefore, the Wave 2 within-household probabilities need to be adjusted to reflect this change. For example, in a household that comprised an age-eligible couple in Wave 1, the within-household selection probability for each individual would have been (essentially) 1 in 2 (or 0.5). Thus, *ceteris paribus*, the weight given individuals selected from such households (those comprising age-eligible couples) in Wave 1 would have included an adjustment (a multiplier of 2) to represent appropriately this within household sampling. In Wave 2, we are including both age-eligible members of such households in our data collection. The weights need to reflect this change. Once the household is selected, there is no further subsampling and therefore no need for the additional factor described earlier for Wave 1. The steps in the construction of the Wave 2 weight_adj variable are given below.

Base weight.—As described in O'Muircheartaigh and colleagues (2009), the selection probability for each case in the NSHAP Wave 1 sample is the product of two factors: (a) the probability of selection of the household by HRS and (b) the subsequent conditional probability of selection of the case into NSHAP survey, given its selection for HRS. The HRS probability incorporates the area stages of

Table 7. Response Rates for NSHAP Wave 2 Partners[a]

Category	Number of eligible cases[b]		Weighted response rate	
	Partners of Wave 1 respondents	All partners	Partners of Wave 1 respondents	All partners
Born 1948 or later	193	211	83.4	82.4
Born 1939–1947	434	471	86.6	84.9
Born 1930–1938	297	316	88.7	86.5
Born 1929 or earlier	132	125	82.3	81.7
Male	406	431	85.4	84.2
Female	643	699	86.1	84.4
All	1,049	1,130	85.8	84.3

Notes. NSHAP = National Social Life, Health, and Aging Project.
[a]The response rates are weighted to take into account the selection probabilities of the NSHAP prime respondents whose partners these are.
[b]The eligible cases are all the partners of Wave 2 respondents who were identified during the Wave 2 interview of the prime.

selection (primary sampling units and segments) and the selection of households for screening within the segment; it includes also other subsampling adjustments carried out by the HRS. The NSHAP probability incorporates the adjustments due to the differential subsampling of nonminority cases in minority segments, the probability of selection of the household for NSHAP, and the probability of selection of the given individual within the household. The base weight is the inverse of the selection probability.

In Wave 2, the first factor—the HRS probability of selection—is unchanged. Most of the second factor is unchanged also. Only the within-household selection probability for eligible individuals differs, and then only if an age-eligible partner was in the household at Wave 1.

If a selected partner was present and age eligible during Wave 1, the within-household selection probability for the primary respondent is now different. We need to consider the pair of respondents—prime and partner—as a single sampling unit. The couple had essentially two different ways of entering the Wave 2 sample; first, by the selection of the prime (which actually happened in Wave 1) and second, had the partner been selected in Wave 1 (which did not happen but had a positive probability of occurrence). The selection probability for the couple in Wave 2 is therefore the sum of the probabilities of selection of the two members of the couple. The base weight for each member of the couple is the inverse of this joint probability of selection (it will be approximately half the base weight for the prime in Wave 1; the household as a whole will still have the same weight). In the great majority of cases, the relationship between the Wave 1 and Wave 2 probabilities is simple. Either there was only one eligible individual in both waves, in which case the probability is unchanged, or there were two in both waves, in which case the Wave 2 probability is roughly twice that in Wave 1.

If a selected partner was not present or was not age eligible in Wave 1, the within-household selection probability (and therefore the base weight) for the primary respondent is unchanged from Wave 1. This base weight will also be the base weight for the partner because they both entered Wave 2 with the probability of selection during Wave 1 of the primary respondent.

Adjustment for eligibility at the time of interviewing.—The base weights need to be adjusted for eligibility because some persons given a base weight were determined to be ineligible. Wave 1 eligibles who died prior to Wave 2 were considered ineligible. Also, some partners screened in became ineligible between screening and interviewing due to leaving the household, being incapacitated, or dying. Ineligible cases are not included in the final data set. No cases were finalized with an unknown eligibility status, and the weights for eligible cases were unchanged in this step.

Adjustment for nonresponse.—Nonresponse of any magnitude threatens the basis of inference from the survey data to the population. We provide an adjustment to the weights to account for nonresponse. All nonresponse adjustments rely on a model that makes assumptions about the nonrespondents. The method we used, which Kalton and Kasprzyk (1986) call sample-based weighting, assumes that, once we control for a few key characteristics, nonrespondents are like respondents. The only variables that can be used to control for nonresponse are those that exist on both the responding and the nonresponding cases. Age provided the greatest discrimination in response rates (see Tables 3, 4, and 7), so we also created cells by crossing four age categories with urbanicity and sample type (primary respondent/partner). In each of the 16 cells formed by these three variables, weights for responding cases were increased by the reciprocal of the cell-level response rate such that the responding cases take on the weight of the nonresponding cases. To the extent that the correspondence between respondents and nonrespondents is closer within these adjustment classes than it is overall, adjusting the weights separately within these classes will improve the validity of our estimates (Kish, 1992). One of the more interesting questions during this adjustment was whether to separate the Wave 1 respondents from the Wave 1 nonrespondents. This issue is described in Pedlow, O'Muircheartaigh, and Schumm (2013). As the Wave 1 nonrespondents were found to be fairly similar to the Wave 1 respondents, we have chosen to release the "no-separation" weights; these will reduce the mean square error, even in the case of minor differences between the two groups.

The couple weight—weight_couple—is appropriate when the unit of analysis is the couple. Such analyses will involve measures derived from both partners' responses. The couples in NSHAP are defined by the age eligibility of at least one partner and their being a couple at the time of the NSHAP Wave 2 fieldwork. The inference population is therefore the population of 2010 couples, at least one of whom is NSHAP age eligible; the NSHAP sample is a probability sample from this population.

We plan to provide in due course a panel weight for those who wish to carry out longitudinal analysis with the data. This weight will apply to respondents who provide data at both waves of interviewing.

Scale adjustment.—As a final step to the weighting procedure, weights are rescaled so that they sum to the total number of completed interviews. This adjustment helps prevent errors in some software packages that calculate the sample size to be the sum of the weights.

Use of Weights and Calculation of Standard Errors

Use of weights.—We recommend that all analyses carried out with NSHAP data incorporate weights because the respondents were not selected with equal probabilities. We suggest that the weights incorporating the adjustment for nonresponse (weight_adj) should be used. Should you wish to carry out your own poststratification adjustments, the unadjusted weight—weight_sel—should be used as the base weight. Otherwise, the estimates will not represent the population and may be subject to serious biases (Kish & Frankel, 1974).

Calculation of standard errors.—Although using weights will ensure that analysts have the right point estimates, the standard errors (and confidence intervals [CIs]) on these estimates will be incorrect unless additional care is taken. To calculate standard errors correctly for NSHAP data, it is necessary to take into account the sample design and the fieldwork outcomes. Importantly, failing to account for the design will lead to serious *underestimation* of standard errors and CIs (Kish, 1965; Lee & Forthofer, 2006; Verma, Scott, & O'Muircheartaigh, 1980). Note that ignoring the design (and underestimating CIs) is the default behavior in most statistical packages, which will lead researchers to conclude that their results are statistically significant when they are not. Three aspects of the sample design can have a substantial effect on standard errors: *stratification*, *clustering*, and *unequal probabilities of selection*. The design effect (deff) summarizes the combined effect of these three influences on the variance of estimates from a sample. The square root of the deff, called the design factor (deft), gives the effect of the design on standard errors. A deff on a given estimate less than one indicates that the estimate from a complex sample is more efficient (has lower variance and standard error) than one from a simple random sample of the same size. A deff greater than 1 indicates that a complex sample gives less efficient estimates.

Stratification tends to reduce the deff, whereas clustering and unequal weights tend to increase it. Different variables within a given survey will have different deff values because some are more highly clustered than other: Variables with high rates of within-cluster homogeneity have a higher deff than those that have low rates of homogeneity (Kreuter & Valliant, 2007).

Tables 8 and 9 present estimates of defts and deffs for a number of variables for Wave 1 and Wave 2, respectively, which were calculated using the nonresponse-adjusted weight (weight_adj). As defts are more relevant to estimation, we discuss the results in those terms; for design purposes, deffs are more relevant as they are inversely related to sample size. A deft of 1.0 indicates that the precision of the estimate is equivalent to that of an unclustered sample; larger defts, indicating that population members are spatially relatively homogeneous on a variable, imply that CIs are wider than they would be for an unclustered sample by a factor equal to deft. The comments below are illustrative of the interpretation of these effects. More considered substantive reasoning should be applied in interpreting analytic results. A comprehensive analysis of design factors for NSHAP is being undertaken by the design team.

Table 8. Design Factors (deft) for Selected Variables in Waves 1 and 2

	Wave 1 deft			Wave 2 deft		
Variable	ALL	ALL	Male	Female	Prime	Partner
COLLEGE	1.97	2.27	1.66	1.96	2.07	1.50
NGRNDCLD	1.54	2.44	1.94	1.85	2.01	1.84
SONS	1.54	1.96	1.86	1.29	1.57	1.69

Table 9. Design Effects (deff) for Selected Variables in Waves 1 and 2

	Wave 1 deff			Wave 2 deff		
Variable	ALL	ALL	Male	Female	Prime	Partner
COLLEGE	3.88	5.17	2.74	3.83	4.28	2.25
NGRNDCLD	2.37	5.96	3.76	3.43	4.05	3.38
SONS	2.38	3.85	3.47	1.67	2.45	2.84

For the education variable, COLLEGE, we see that the deft for Wave 2 is slightly higher than for Wave 1. This is probably due to the sample size for Wave 2 being somewhat larger; the deff is a function of both the degree of homogeneity and the number of respondents from each cluster. Further examination shows that it is for women that this increase in deff has occurred. (Among Wave 1 respondents, there were 1,455 men and 1,545 women; in Wave 2, there were 1,598 men and 1,727 women.) In general, for education, there is a high degree of within-cluster homogeneity; people who live in the same neighborhood tend to have similar levels of education. This clustering in the variable means that our design is much less efficient at estimating this characteristic than an unclustered sample would be. CIs on estimates of the proportion of the NSHAP population that has a college education will be about twice as large (multiplier for the standard error = 2.27) as those from a simple random sample of the same size.

The example here deals with the estimation of a sample mean. Correctly estimating standard errors requires passing stratum, cluster, and weight variables into appropriate statistical software. Most packages offer special routines for this kind of estimation: Stata's SVY commands, R's survey package, and SAS's PROC SURVEYMEANS, PROC SURVEYFREQ, etc. The NSHAP data file includes stratum and cluster identifiers as well as the weights so that these can be passed into the software routine to produce appropriate standard errors.

Conclusions

The second wave of NSHAP was designed to increase the scientific value of the Wave 1 data set by revisiting sample members 5 years after their initial interviews and augmenting this sample where possible.

A conditional panel response rate of 89% was achieved, corresponding to an annualized rate of 98%, which exceeded expectations and compares favorably with conditional response rates for other major U.S. national longitudinal surveys; the oldest sample members those aged 80 and older had significantly lower response rates. Nonhostile nonrespondents from Wave 1 were approached again in Wave 2 and yielded a response rate of 26%. This combination of response rates produced an unconditional response rate for Wave 2 of 74%, only two percentage points below the response rate for Wave 1.

One respondent per household was selected in the design for Wave 1 of NSHAP (the prime respondents). In Wave 2, coresident spouses or romantic partners of Wave 1 sample members were added; this addition has considerable analytic and economic advantages, but there was a concern that this inclusion might have a deleterious effect either on the response rates of the primes or on the quality of the responses. The Wave 2 fieldwork incorporated an experiment to test this; no negative effect on response rates was observed. A future paper will examine the possible effect on the responses. The response rates for the partners, conditional on their being identified, was 84% (86% for the partners of Wave 1 respondents).

Wave 2 of NSHAP has maintained the high level of representativeness produced by Wave 1; by returning to nonrespondents from Wave 1, the degree of attrition has been greatly ameliorated, The sample has been augmented by the addition of partners, and this increase in the sample size has preserved the inferential strength of the sample even as the sample ages, and mortality takes its toll.

Key Points

- NSHAP Wave 2 preserves the exceptional level of representation obtained in Wave 1; the unconditional response rate for the panel members was 74%.
- Nonrespondents from Wave 1 were again approached for interview in Wave 2 to compensate for the expected attrition from Wave 1; 26% of these responded, reducing the net attrition between waves by more than three quarters.
- The scope of the study was expanded by adding coresident spouses or romantic partners; the conditional response rate for partners was 84%.
- Early analyses (based on a controlled experiment built into the study) suggest that the inclusion of partners did not have a deleterious effect either on the response rates or on the data quality of the original panel members.
- The methodology of NSHAP Wave 2 achieved its objectives of consolidating the panel, recruiting partners, and converting a significant proportion of Wave 1 nonrespondents.

Funding

The National Social Life, Health, and Aging Project is supported by the National Institutes of Health, including the National Institute on Aging (R37AG030481; R01AG033903), the Office of Women's Health Research, the Office of AIDS Research, and the Office of Behavioral and Social Sciences Research (R01AG021487), and by NORC which was responsible for the data collection. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, or NORC.

Correspondence

Correspondence should be addressed to Colm O'Muircheartaigh, PhD, University of Chicago Harris School and NORC, 1155 East 60th Street, Chicago, IL 60637. E-mail: caomuirc@uchicago.edu.

References

Harter, R., Eckman, S., English, N., & O'Muircheartaigh, C. (2010). Applied sampling for large-scale multi-stage area probability designs. In P. Marsden & J. Wright (Eds.), *Handbook of survey research* (2nd ed., pp. 169–198). Bingley, UK: Emerald Group Publishing Limited.

Jaszczak, A., O'Doherty, K., Colicchia, M., Satorius, M., McPhillips, J., Czaplewski, M., . . . Smith, S. (2014). Continuity and innovation in the data collection protocols of the second wave of the National Social Life, Health, and Aging Project [under review]. *Journals of Gerontology, Psychological Sciences and Social Sciences*.

Kalton, G., & Kasprzyk, D. (1986). The treatment of missing survey data. *Survey Methodology*, *12*, 1–16.

Kish, L. (1965). *Survey sampling*. New York, NY: Wiley.

Kish, L. (1992). Weighting of unequal pi. *Journal of Official Statistics*, *8*, 183–200.

Kish, L., & Frankel, M. (1974). Inference from complex samples. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *36*, 1–22.

Kreuter, F., & Valliant, R. (2007). A survey on survey statistics: What is done and can be done in Stata. *Stata Journal*, *7*, 1–21.

Lee, E. S., & Forthofer, R. N. (2006). *Analyzing complex survey data* (2nd ed.). Thousand Oaks, CA: Sage.

O'Muircheartaigh, C., Eckman, S., & Smith, S. (2009). Statistical design and estimation for the national social life, health, and aging project. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, *64B*, i12–i19. doi:10.1093/geronb/gbp045

Pedlow, S., O'Muircheartaigh, C., & Schumm, P. (2013). NSHAP's Wave 2 nonresponse weight adjustment with some Wave 1 nonresponses. 2013 Proceedings of the American Statistical Association, Survey Research Methods Section [CD-ROM]. Alexandria, VA: American Statistical Association.

Schoeni, R. F., Stafford, F., McGonagle, K. A., & Andreski, P. (2013). Response rates in national panel surveys. *The Annals of the American Academy of Political and Social Science*, *645*, 60–87. doi:10.1177/0002716212456363

Verma, V. K., Scott, C., & O'Muircheartaigh, C. (1980). Sample designs and sampling errors for the world fertility survey (with discussion). *Journal of the Royal Statistical Society, Series A: Statistics in Society*, *143*, 431–473.