

Missing Data in Wave 2 of NSHAP: Prevalence, Predictors, and Recommended Treatment

Louise C. Hawley,¹ Masha Kocherginsky,² Jaclyn Wong,³ Juyeon Kim,⁴ and Kathleen A. Cagney³

¹Academic Research Centers, NORC at the University of Chicago, Illinois.

²Department of Health Studies and

³Department of Sociology, University of Chicago, Illinois.

⁴Department of Sociology, National University of Singapore.

Objectives. This report seeks to inform National Social Life, Health, and Aging Project (NSHAP) data users of the prevalence and predictors of missing data in the in-person interview (CAPI) and leave-behind questionnaire (LBQ) in Wave 2 of NSHAP, and methods to handle missingness.

Method. Missingness is quantified at the unit and item levels separately for CAPI and LBQ data, and at the item level is assessed within domains of conceptually related variables. Logistic and negative binomial regression analyses are used to model predictors of unit- and item-level nonresponse, respectively.

Results. Unit-level nonresponse on the CAPI was 10.6% of those who responded at Wave 1, and LBQ nonresponse was 11.37% of those who completed the Wave 2 CAPI component. CAPI item-level missingness was less than 1% of items for most domains but 7.1% in the Employment and Finances domain. LBQ item-level missingness was 5% across domains but 8.3% in the Attitudes domain. Missingness was predicted by characteristics of the sample and features of the study design.

Discussion. Multiple imputation is recommended to handle unit- and item-level missingness and can be readily and flexibly conducted with multiple imputation by chained equations, inverse probability weighting, and in some instances, full-information maximum-likelihood methods.

Key Words: Full-information maximum-likelihood—Inverse probability weighting—Missing data—Multiple imputation by chained equations.

MISSING data are unavoidable in large-scale survey research. Careful a priori selection of data collection methodologies to minimize missing data is important, but some degree of “missingness” will persist. For instance, in the National Social Life, Health, and Aging Project (NSHAP), respondents could refuse or be unable to answer some or all of the items administered during the computer-assisted personal interview (CAPI). CAPI data could be lost through human error and technical failures. Self-report questionnaire data (i.e., the “Leave-Behind Questionnaire,” or LBQ) could be missing entirely if the forms were not returned or were lost in the mail, and the forms that were returned could be missing data on individual items or portions of the questionnaire. Missing data can be disconcerting but are not devastating if understood and appropriately handled. This article seeks to inform NSHAP data users of the prevalence of missing CAPI and LBQ data in Wave 2 of NSHAP, predictors that help to explain missingness, and methods to handle these missing data, at least in cross-sectional analyses involving Wave 2 (or Wave 1) data. Those doing longitudinal analyses should be aware that there is information in the data set on the disposition of all Wave 1 respondents in Wave 2 (i.e., death, illness, and nonresponse), and this, together with how these three sources

of nonresponse are related to respondent characteristics at Wave 1, should be considered by anyone doing longitudinal analyses.

Data can be missing at the unit level and the item level. Unit nonresponse refers to the complete absence of data from a respondent, whereas item nonresponse refers to missing answers to only some questions. Missing data are important to survey researchers because, depending on the type of missing data and how they are handled, results may be biased and not generalizable to the population of interest. Statistical methods that are now readily available in many standard statistical software packages can be applied to handle missing data appropriately.

How missing data are handled in analyses depends on the missingness mechanism. Data can be missing completely at random (MCAR), conditionally missing at random (MAR), or missing not at random (MNAR), terms introduced by Little and Rubin (1987). Data are MCAR if the probability of missingness is independent of both the observed and unobserved variables (Little & Rubin, 2002). For instance, any of the randomized modules within NSHAP yield MCAR data among those not selected. Complete case analysis (i.e., casewise deletion) of MCAR data is valid because these cases can be considered a random sample of all cases, but

the analysis could be inefficient because of reduced sample size. Most of the time, however, missing data are related to at least some characteristics of the sample, the data collection method, or other features of the study. Data are MAR when the probability of missingness does not depend on the unobserved data but can be fully explained by the observed variables. That is, after one controls for observed variables, the data can be considered MCAR. For MNAR data, the probability of missingness depends on unmeasured characteristics of the sample even after conditioning on the observed data. MNAR are a particular concern in longitudinal studies in which respondent attrition is related to their initial status on the outcome of interest (e.g., those too ill to be interviewed in Wave 2). Most missing data fall along a continuum between MAR and MNAR (Graham, 2009).

The goals of statistical analysis in the presence of missing data are to minimize bias in parameter estimates, maximize the use of available data, and obtain efficient and accurate assessments of the statistical uncertainty of the estimated quantities (Allison, 2001). There are three major approaches to dealing with missing data, and any of them can yield good results if performed properly. In survey research, weighting and multiple imputation (Little and Rubin, 1987) are the most common approaches for handling missingness. Multiple imputation by chained equations (MICE) is a particularly appealing imputation approach due to its flexibility and availability in standard statistical software (e.g., Stata and R). Recent developments in doubly robust inverse probability weighting (IPW) methods have revealed this to be an attractive approach that can be competitive with multiple imputation (Carpenter, Kenward, & Vansteelandt, 2006; Seaman & White, 2013). A full-information maximum-likelihood (FIML)-based approach, although computationally more challenging than multiple imputation and IPW methods, is particularly appropriate in situations in which missing data mechanisms are unknown and can perform as well or better than multiple imputation under certain circumstances (Larsen, 2011). For a detailed review of the statistical methodology for missing data, we refer the reader to the study by Little and Rubin (2002).

METHOD

NSHAP is a nationally representative study of community residing adults born between 1920–1947 (i.e., age eligible respondents). At the time of the Wave 2 interview (2010–11), they were 62–91 years of age, and included an oversampling of African Americans, Hispanics, and the oldest old. Wave 2 was conducted in 2010–2011 and included cohabiting spouses and romantic partners. NSHAP uses a complex, multi-stage area probability sample with poststratification. Sample design details are reported by O’Muircheartaigh, Eckman, and Smith (2009) and O’Muircheartaigh, English, Pedlow, and Kwok (2014). Data are publically 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. 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.).

We assess missingness in Wave 2 of NSHAP for each level of missingness (unit level and item level) within each mode of administration (CAPI, LBQ). We categorize item-level missingness within domains of potential interest to data users. Within each domain, we evaluate missingness on responses to primary questions asked of all respondents, and not on conditional branch questions that, by definition, are asked of only a subset of respondents. The domains correspond to sections within the table of contents in the booklet version of the CAPI questionnaire (i.e., *nshap_w2_inperson_questionnaire.pdf*) and the LBQ (i.e., *nshap_w2_leavebehind.pdf*).

In the CAPI, the seven domains are (a) *Basic background information* (gender, age, education, and race–ethnicity); (b) *Social context* (social support from family and friends); (c) *Physical health* (self-rated health, sensory function, access to health care, chronic diseases, functional health, and health-related behaviors); (d) *Children and grandchildren* (number of living children and grandchildren); (e) *Mental health* (happiness and depressive symptoms); (f) *Employment and finances* (employment, income, and assets), and (g) *Religion* (religious preference and service attendance).

In the LBQ, the 10 domains are (a) *Childhood background* (country of origin, quality of family life in childhood, parental education, family’s socioeconomic status, childhood health status, and experienced or witnessed violent event during childhood); (b) *Social relationships and activities* (frequency and quality of interactions with friends and relatives); (c) *Bereavement* (death of close other in the last 5 years); (d) *Neighborhood* (residential duration, frequency of interactions among neighbors to visit/do favors/ask advice; ratings of closeness, similarity to, and trust of people in neighborhood); (e) *Caregiving* (assists an adult needing help); (f) *Attitudes* (toward physical touch, frequency and importance of touching, frequency and importance of sex, and quality of sex life); (g) *Thoughts and feelings* (personality traits, loneliness, anxiety, and perceived stress); (h) *Health* (pain, falling/breaking bones, having had surgery/skin disease/gum disease, and frequency and duration of daily naps); (i) *Fertility* (number of children parented and intended, number of biologically related grandchildren, and age at first pregnancy or fathered child); and (j) *Background* (health care coverage and sources; role of religious beliefs in daily life; military service history; household income relative

to others; having experienced unwanted sexual advances; having been a victim of violent crime in the past 2 years).

Predictors of Missingness

Sociodemographic variables.—Age was further categorized into three groups labeled: 62–69 years, 70–79 years, and 80–90 years old (including two participants who were 91 as an artifact of interview scheduling). *Gender* contrasted males and females. *Race–ethnicity* contrasted five groups: white or Caucasian, black or African American, American Indian, Asian or Pacific Islander, and other (11 nonidentified cases). *Education* contrasted four groups: less than high school, high school or equivalent, vocational certification or some college, and bachelor’s degree or more. For *household income*, we employed the bracketed categories that NSHAP used to successfully reduce nonresponse to the open-ended question about income described for HRS (Juster & Smith, 1997; St. Clair et al., 2011) and implemented in NSHAP. We therefore compared missingness in four groups: <\$25K, \$25–49K, \$50–99K, and ≥\$100K. Similarly, we used bracketed categories of *household assets* to compare missingness in five groups: <\$10K, \$10–49K, \$50–99K, \$100–499K, and ≥\$500K. For *marital status*, we compared six groups: married, living with a partner, separated, divorced, widowed, and never married. Finally, *living arrangements* compared partnered/married and unpartnered/unmarried respondents living with someone or living alone.

Health status and cognitive functioning.—Single items for *self-rated physical health* and *self-rated mental health* were each rated on a 5-point scale: poor, fair, good, very good, and excellent. Values of a composite *cognitive function* measure (Shega, et al., 2014) ranged from 0 to 20, where higher values signify better cognitive function.

Respondent type.—In Wave 2 of NSHAP, considerable effort was deployed to convert respondents who refused to participate in Wave 1 and convince them to participate in Wave 2. These efforts recovered 161 respondents (O’Muircheartaigh et al., 2014). However, a recent review showed that converted initial refusers are subsequently less likely to provide complete data than initial cooperators (Yan & Curtin, 2010). We compared Wave 2 missingness in Wave 1 initial nonresponders, Wave 1 initial responders, and partners whose initial response was in Wave 2. Inclusion of the latter group allowed us to determine whether partners’ initial response rate (Wave 2) differed from primes’ initial response rate (Wave 1).

Biomeasure path.—In Wave 2 of NSHAP, respondents were randomly assigned to one of six interview “paths” that differed in the content of the biomeasure component. Saliva was collected in four paths (2, 4, 5, and 6); the smell test was administered in a different set of four paths (1, 3, 5, and 6); and the Actiwatch protocol was administered in only two

paths (2 and 6). These paths are summarized in the study by Jaszczak and coworkers (2014, Figure 1). We tested the possibility that differential respondent burden among these paths would influence the likelihood of LBQ return. Unlike other paths, those assigned to an Actiwatch path were asked to participate in a 3-day at-home protocol that involved wearing an accelerometer and completing a daily sleep log (Lauderdale et al., 2014). Assignment to the Actiwatch paths would also have entailed repeated contact with NSHAP staff, initially to confirm their willingness to participate in the Actiwatch module and schedule a time to have the watch delivered and, subsequently, for those who failed to return the watch, contacts (often several) to remind them to do so. Differences in respondent burden may have influenced the likelihood of participants completing and returning the LBQ, and we therefore compared missingness among the six biomeasure paths.

Statistical Analyses

Missingness prevalence is reported overall and by age categories within gender to correspond to the targeted sample cells in Wave 1. In addition, missingness is reported by predictor category for each predictor of missingness.

Logistic regression analyses are used to model predictors of unit-level nonresponse. Negative binomial regression analyses (to accommodate count data outcomes such as the number of missing items in a domain) are used to model predictors of item-level missingness in select CAPI and LBQ domains. Statistically significant ($p < .05$) global F tests are followed up by examining regression coefficients to determine the source of significant associations between predictors and missingness.

All analyses use the survey design (clustered and stratified) and a set of weights (*weight_adj*) that represent the inverse probability of selection and are adjusted for CAPI nonresponse based on age, urbanicity, and sample type (primary respondent vs. partner; O’Muircheartaigh et al., 2014). In addition, analyses are limited to age-eligible respondents from 62 to 90 years old (the age range corresponding to the targeted sample in Wave 1). These adjustments permit making inferences about rates of missingness in the population of adults aged 62–90 years.

RESULTS

Unit-Level Missingness

Prevalence.—Among those who responded at Wave 1, 89.4% responded at Wave 2 (O’Muircheartaigh et al., 2014), and all except one partial case completed the entire CAPI. Correlates of and adjustments for Wave 2 unit-level nonresponse have been reported elsewhere (O’Muircheartaigh et al., 2014); these are not discussed further here. In the case of the LBQ, data were missing entirely for some respondents who completed the Wave 2 CAPI. The rate of LBQ survey nonresponse, as gauged by

failure of respondents to return the LBQ in the mail, was 11.37% of those who participated in the in-person interview. Rates of LBQ nonreturn are shown in Table 1 for each of the age–gender categories, where it can be seen that the rate of nonreturn across age groups ranges from approximately 9% to 15%. An overall nonreturn rate of 11.37% is highly comparable to that observed in the Health & Retirement Study, where 10% (in 2006) and 11% (in 2008) of those who completed the enhanced face-to-face interview failed to return the leave-behind questionnaire (Smith et al., 2013). Moreover, the NSHAP rate of nonreturn is only half the average percentage of missing cases

in studies reviewed by Eekhout, De Boer, Twisk, De Vet, and Heymans (2012). These researchers extracted information on missing data from 262 eligible studies published in the *American Journal of Epidemiology*, *Epidemiology*, and the *International Journal of Epidemiology* and found that an average of 26% of cases were missing all data (range <1%–82%). Notably, however, this estimate was based on a subset of the 262 studies because only 32% of these studies reported missingness at the unit or case level, a reporting deficit that may signal unawareness of the impact of missing data on study results (Eekhout et al., 2012).

Table 1. Mean Percent Leave-Behind Questionnaire Nonreturn Rate (SD) by Age–Gender Categories

	62–69 years	70–79 years	80–90 years
Males	12.64 (30.6)	11.70 (33.9)	10.40 (33.4)
Females	10.48 (29.3)	8.96 (29.8)	15.07 (37.9)

Predictors of unit-level LBQ missingness.—Nonreturn rates are documented in the first data column of Table 2 for each predictor category. Logistic regression coefficients and probability levels for each predictor are displayed in the second and third data columns of Table 2 and are discussed further subsequently.

Table 2. Leave-Behind Questionnaire (LBQ) Nonreturn Rates by Predictor Category and Logistic Regression Coefficients Predicting the Log Odds of LBQ Nonreturn

Predictor	% of LBQ's not returned (linearized SE)	B (linearized SE)	p
Age group			
62–69 years (ref)	11.52 (1.25)	—	
70–79 years	10.29 (0.97)	–0.13 (0.17)	.455
80–90 years	12.99 (1.56)	0.14 (0.18)	.461
Gender			
Male (ref)	11.87 (1.16)	—	
Female	10.92 (0.86)	–0.09 (0.14)	.505
Race/Ethnicity*			
White/Caucasian (ref)	9.52 (0.72)	—	
Black/African American	27.64 (3.87)	1.29 (0.22)	<.001
American Indian	14.99 (9.35)	0.52 (0.76)	.492
Asian or Pacific Islander	(Empty)	(Empty)	—
Other	16.14 (2.88)	0.60 (0.23)	.01
Education*			
Less than high school (ref)	16.85 (1.56)	—	
High school/Equivalent	11.69 (1.14)	–0.43 (0.16)	.01
Vocational certification	9.70 (1.13)	–0.64 (0.17)	<.001
Bachelor's degree or more	9.72 (1.51)	–0.63 (0.20)	.002
Household income			
Less than \$25,000 (ref)	14.74 (1.72)	—	
\$25,000–49,000	10.01 (1.52)	–0.44 (0.23)	.064
\$50,000–99,000	9.51 (1.48)	–0.50 (0.22)	.027
\$100,000+	8.65 (1.71)	–0.60 (0.26)	.023
Household assets*			
Less than \$10,000 (ref)	20.32 (3.11)	—	
\$10,000–49,000	11.66 (1.91)	–0.66 (0.28)	.022
\$50,000–99,000	13.26 (2.56)	–0.51 (0.23)	.032
\$100,000–499,000	9.10 (0.99)	–0.94 (0.23)	<.001
\$500,000+	8.84 (1.47)	–0.97 (0.26)	.001
Marital status*			
Married (ref)	9.29 (0.95)	—	
Living with a partner	12.51 (4.90)	0.33 (0.43)	.44
Separated	19.47 (8.83)	0.86 (0.60)	.157
Divorced	16.45 (3.40)	0.65 (0.27)	.02
Widowed	13.84 (1.53)	0.45 (0.18)	.013
Never married	18.81 (5.07)	0.82 (0.36)	.029

(Table 2 continues)

Table 2 (continued)

Predictor	% of LBQ's not returned (linearized SE)	B (linearized SE)	p
Living arrangements*			
Partnered, living alone (ref)	21.41 (5.82)	—	
Partnered, living with someone	9.54 (1.01)	-0.95 (0.36)	.011
Unpartnered, living alone	11.80 (1.26)	-0.71 (0.38)	.065
Unpartnered, living with someone	19.51 (3.02)	-0.12 (0.40)	.773
Self-rated physical health*			
Poor	21.86 (3.34)	—	
Fair	14.71 (1.64)	-0.48 (0.23)	.039
Good	9.67 (0.91)	-0.96 (0.23)	<.001
Very good	9.31 (1.30)	-1.00 (0.22)	<.001
Excellent	10.78 (2.17)	-0.84 (0.31)	.009
Self-rated mental health*			
Poor	17.51 (5.81)	—	
Fair	17.16 (2.25)	-0.02 (0.46)	.958
Good	11.35 (1.36)	-0.50 (0.42)	.236
Very good	8.66 (1.03)	-0.81 (0.41)	.053
Excellent	12.86 (1.71)	-0.36 (0.45)	.425
Cognitive functioning*			
Continuous measure (0–20)		-0.11 (0.02)	<.001
Respondent type*			
Wave 1 respondent (ref)	11.38 (0.84)	—	
Wave 1 nonrespondent	23.96 (4.13)	0.90 (0.25)	.001
Partner	7.86 (1.13)	-0.41 (0.18)	.026
Biomeasure path*			
1	9.49 (1.60)	—	
2	16.94 (2.54)	0.67 (0.28)	.021
3	8.21 (1.33)	-0.16 (0.27)	.558
4	9.12 (1.53)	-0.04 (0.28)	.875
5	9.66 (1.42)	0.02 (0.23)	.932
6	14.81 (2.51)	0.51 (0.30)	.094

Note. *Global F test is significant, $p < .05$.

Sociodemographic factors. Global F tests revealed that LBQ return rates differed with race–ethnicity, education, household assets, marital status, and living arrangements, $ps < .05$. Age, gender, and household income were not related to LBQ return rates, $ps > .06$. For *race–ethnicity*, logistic regression coefficients showed that the nonreturn rate was significantly higher in the black/African American group and the “other” racial-ethnic group than in the white/Caucasian group, $ps \leq .01$ (see Table 2). For *education*, the nonreturn rate in those with less than a high school education was significantly greater than in each of the other education categories. In general, the nonreturn rate diminished with increased education. For *household assets*, significant differences were noted between the lowest asset category (<\$10K) and each of the higher asset categories. Overall, the nonreturn rate tended to decrease with increasing assets. For *marital status*, the nonreturn rate was significantly lower in the married group than in the divorced, widowed, and never married groups. The separated group had the highest rate of nonreturn (19.47%, see Table 2), but the statistical test was unreliable. If spousal influence (e.g., reminders to complete the LBQ) contributed to enhanced return rates in the married group relative to the other groups, then spouses with a partner in the study should exhibit lower nonreturn rates than spouses without a partner

in the study. A post hoc comparison confirmed expectations: the nonreturn rate for spouses with a partner in the study was 8.46% (linearized $SE = 1.21$), whereas the nonreturn rate for primes whose partner was not in the study was 14.47% (linearized $SE = 1.89$), $p < .001$. Finally, for *living arrangements*, the nonreturn rate was significantly higher in respondents with a romantic noncohabiting partner than in partnered respondents living with someone.

Health. Global F tests revealed that LBQ return rates differed with each of the health variables, $ps < .01$. For *physical health*, nonreturn rates were significantly higher in the poorest health group than in each other health group, as shown in Table 2. In general, rates of nonreturn decreased with increased health. For *mental health*, the nonreturn rate differed in a less systematic fashion. Post hoc comparisons revealed a significant difference only between the nonreturn rate of the fair and very good groups, p (Scheffé adjustment) = .018. For *cognitive functioning*, the log odds of LBQ nonreturn diminished with each unit increase in cognitive function (see Table 2).

Respondent type. A significant global F test, $p < .001$, arose from higher rates of nonreturn in Wave 1

nonresponders than Wave 1 responders, and significantly lower rates of nonreturn among partners than Wave 1 responders (see Table 2). Follow-up analyses revealed that partners exhibited a lower nonreturn rate than the prime respondents who were their spouses, $p = .041$, suggesting that novelty contributed to higher return rates in partners newly introduced to the study in Wave 2.

Biomeasure path. A global F test indicating significant differences in LBQ nonreturn among the six biomeasure paths, $p = .011$, arose from a higher rate of nonreturn in Path 2 (an Actiwatch path) than Path 1, and a sizeable but nonsignificantly higher rate of nonreturn in Path 6 (an Actiwatch path) than Path 1 (see Table 2). Post hoc comparisons revealed that Path 2 also had a significantly higher rate of nonreturn than Paths 3, 4, and 5, ps (Scheffé adjustment) $< .02$. Path 6 had a significantly higher rate of nonreturn than Paths 3 and 4, ps (Scheffé adjustment) $< .05$. Overall, the pattern of results supported the hypothesis that the Actiwatch protocol in Paths 2 and 6 had a modestly adverse effect on LBQ return rates, with nonreturn rates exceeding the overall average rate of 11.37% by 3–5 percentage points.

This indication of lower return rates among respondents in the Actiwatch paths prompted the question whether response likelihood differed more systemically such that the respondents in these two paths were also less likely to comply with in-person portions of the survey and the biomeasures than in the other paths. We tested whether those who failed to return the LBQ were also more likely to have missing data on CAPI questions about household income, household assets, and sexuality (specifically, “how often do you think about sex?”). Results for income and assets were nonsignificant, $ps > .2$, indicating no biomeasure path differences in missingness on these two items. Missingness on the sexuality item showed only a modest tendency to differ among biomeasure paths, $p = .054$, and post hoc comparisons revealed no significant path differences in sexuality item missingness. Using systolic blood pressure (SBP) as a proxy for other biomeasures obtained in each path, we found no difference in SBP missingness among biomeasure paths, $p = .245$. It seems, therefore, that random assignment to biomeasure path was successful; respondents assigned to Paths 2 and 6 did not differ systematically from those assigned to the other paths. Instead, the Actiwatch protocol seems to have played a role in diminishing the LBQ return rate.

Treatment.—Unit nonresponse is usually handled by adjusting sampling weights (Little, 1988). Weights are applied to each respondent’s record to reduce bias in survey estimates and more accurately estimate effects in the target population. Weights compensate for nonresponse by being adjusted upward for respondents who represent nonrespondents. The adjustment is usually made by modeling the probability of being a respondent as a function of data available on all sampled units. This adjustment is implemented in the

respondent-level weights provided in Wave 2 of NSHAP, but these weights do not adjust for the LBQ nonresponse.

Multiple imputation is an appropriate and effective method to handle unit nonresponse, LBQ nonresponse in the case of NSHAP’s Wave 2. The key steps in multiple imputation are to (a) generate multiple complete case data sets by repeatedly drawing values from the posterior predictive distribution of the missing data, (b) analyze each complete data set using standard statistical methods, and (c) combine estimates from each data set to obtain the overall parameter and variance estimates. Multiple plausible values are imputed in order to reflect the additional uncertainty. Multiple imputation can increase estimation efficiency if the data are MCAR and provides unbiased estimates when the data are MAR and the cause of missingness is taken into account. The foregoing analyses indicated that LBQ nonresponse is at least partially explained by race–ethnicity, education, household assets, marital status, living arrangements, self-rated physical and mental health, cognitive functioning, respondent type, and biomeasure path. These variables should therefore be considered candidates for inclusion in the imputation models together with all variables in the analysis model, including the dependent variable (Graham, 2009).

How the data are imputed is key for obtaining valid statistical results—the imputation model should preserve the joint distribution of the variables. The recently proposed MICE approach (van Buuren, 2007) is a practical approach that can accommodate continuous, binary, categorical, and ordinal variable types, and allows simultaneous specification of a different imputation model for each variable. MICE is particularly appealing because of its flexibility and its availability in standard statistical software, for example, “mi impute chained” in Stata. Royston and White (2011) discuss practical aspects and provide guidance on the use of MICE using examples in Stata.

Item-Level Missingness

Prevalence in CAPI.—Rates of item-level missingness in the CAPI are low overall. Among the 105 items included in these analyses, only 1.21 items were missing data on average ($SD = 1.80$, range = 0–45). Missingness rates are reported by age–gender categories in Table 3. Item missingness was skewed; 46% of respondents were missing zero items, 95% of respondents were missing fewer than 5 items, and less than 0.1% of respondents were missing more than 20 items. The mean number of missing items represents a missingness rate of 1.15% ($SD = 1.71$, range = 0%–43%) that compares highly favorably with rates in 262 epidemiological studies

Table 3. Mean Percent Computer-Assisted Personal Interview Item-Level Missingness (SD) by Age–Gender Categories

	62–69 years	70–79 years	80–90 years
Males	0.81 (1.17)	1.16 (1.93)	1.42 (2.41)
Females	1.05 (1.48)	1.18 (1.50)	1.69 (2.32)

published in 2010 and summarized by Eekhout and coworkers (2012). Although only 5% of the 262 studies provided information on item-level missingness, Eekhout and coworkers (2012) found that item-level missingness averaged 11% (range = 1%–44%) in those 13 studies.

Among the domains of CAPI variables, the highest rate of missing items was in the Employment and Finances domain ($M = 7.1\%$, $SD = 9.7$), followed by “Access to Health Care” in the Physical Health domain ($M = 3.8\%$, $SD = 12.1$), and the Social Context domain ($M = 1.9\%$, $SD = 7.7$). All other domains and subsections of the Physical Health domain were missing data on less than 1% of the items.

Predictors of item-level missingness in CAPI.—Because item-level missingness was low overall, we chose to model missingness in the domain with the highest rate of missingness, a domain that encompasses Employment and Finances, variables in demand by researchers in diverse fields. We employed the same predictors as were used for LBQ nonreturn, with the exception of household income and household assets that we did not use as predictors because they are part of the outcome of interest. Missingness rates by predictor category are displayed in the first data column of Table 4.

Table 4. Mean Percent Items Missing in the Computer-Assisted Personal Interview Employment and Finances Domain (7 Variables) and Negative Binomial Regression Coefficients Estimating the Log Number of Items Missing by Predictor Category

Predictor	% of items missing (linearized SE)	B (linearized SE)	P
Age group*			
62–69 years (ref)	5.73 (0.56)	—	
70–79 years	7.23 (0.39)	0.23 (0.08)	.006
80–90 years	9.70 (0.55)	0.53 (0.08)	<.001
Gender*			
Male (ref)	5.49 (0.46)	—	
Female	8.51 (0.45)	0.44 (0.08)	<.001
Race/Ethnicity*			
White/Caucasian (ref)	6.81 (0.38)	—	
Black/African American	9.46 (0.94)	0.33 (0.08)	<.001
American Indian	9.33 (2.30)	0.31 (0.22)	.150
Asian or Pacific Islander	8.45 (1.28)	0.22 (0.14)	.137
Other	5.65 (0.94)	–0.19 (0.17)	.264
Education*			
Less than high school (ref)	10.06 (0.90)	—	
High school/Equivalent	8.25 (0.58)	–0.20 (0.09)	.037
Vocational certification	6.28 (0.41)	–0.47 (0.07)	<.001
Bachelor’s degree or more	5.05 (0.53)	–0.69 (0.12)	<.001
Marital status*			
Married (ref)	7.08 (0.44)	—	
Living with a partner	7.34 (1.21)	0.04 (0.17)	.835
Separated	3.51 (1.14)	–0.70 (0.33)	.041
Divorced	5.04 (0.59)	–0.34 (0.10)	.002
Widowed	8.22 (0.68)	–0.15 (0.07)	.033
Never married	5.81 (1.07)	–0.20 (0.18)	.282
Living arrangements			
Partnered, living alone (ref)	5.55 (1.16)	—	
Partnered, living with someone	7.09 (4.30)	0.24 (0.19)	.201
Unpartnered, living alone	7.05 (5.86)	0.24 (0.20)	.229
Unpartnered, living with someone	7.64 (7.20)	0.32 (0.21)	.134
Self-rated physical health			
Poor	7.29 (1.28)	—	
Fair	8.04 (0.65)	0.10 (0.16)	.546
Good	7.78 (0.49)	0.06 (0.15)	.678
Very good	6.15 (0.50)	–0.17 (0.15)	.279
Excellent	5.83 (0.56)	–0.22 (0.19)	.246
Self-rated mental health*			
Poor	11.55 (1.99)	—	
Fair	8.68 (0.84)	–0.29 (0.22)	.194
Good	7.21 (0.61)	–0.47 (0.19)	.015
Very good	6.69 (0.54)	–0.55 (0.20)	.009
Excellent	6.42 (0.41)	–0.59 (0.18)	.002
Cognitive functioning*			
Continuous measure	—	–0.07 (0.01)	<.001

(Table 4 continues)

Table 4 (continued)

Predictor	% of items missing (linearized SE)	B (linearized SE)	P
Respondent type*			
Wave 1 respondent (ref)	6.79 (0.42)	—	
Wave 1 nonrespondent	8.59 (0.78)	0.24 (0.09)	.012
Partner	7.81 (0.60)	0.14 (0.05)	.008
Biomeasure path			
1	6.89 (0.49)	—	
2	7.25 (0.69)	0.05 (0.10)	.607
3	7.28 (0.50)	0.06 (0.09)	.551
4	6.73 (0.92)	-0.02 (0.14)	.869
5	7.72 (0.52)	0.11 (0.09)	.230
6	6.51 (0.64)	-0.06 (0.09)	.547

Note. *Global F test is significant, $p < .05$.

Sociodemographic factors. Negative binomial regression models revealed statistically significant F tests indicating that missingness in the Employment and Finances domain was associated with age, gender, race–ethnicity, education, and marital status, but not with living arrangements. Table 4 displays negative binomial regression coefficients for each predictor. For *age*, item-level missingness increased with each increment in age group. *Females* showed a higher rate of missingness than males, perhaps reflecting that women in this age group are less familiar than men with the household financial details. For *race–ethnicity*, item-level missingness was greater among blacks/African Americans than whites/Caucasians. For *education*, item-level missingness decreased with each increment in educational category. For *marital status*, item-level missingness was greater in married than in separated, divorced, and widowed respondents, perhaps reflecting less familiarity with household financial details among married women, in particular, than among single individuals who are solely responsible for their finances. In addition, post hoc comparisons revealed that widowed respondents had higher item-level missingness than their divorced counterparts, p (Scheffé adjustment) = .015. No other comparisons were statistically significant, $ps > .05$.

Health status and cognitive functioning. Negative binomial regression models revealed statistically significant F tests indicating that missingness rates in the Employment and Finances domain were associated with self-rated physical health, self-rated mental health, and cognitive functioning. For *self-rated physical health*, neither the regression coefficients (Table 4) nor post hoc comparisons revealed any significant group differences, ps (Scheffé adjustments) $> .06$. For *self-rated mental health*, item-level missingness decreased with each increment in mental health category. *Cognitive functioning* showed an inverse association with item-level missingness; as shown in Table 4, each one-unit increase in cognitive functioning was associated with a decrease in the log number of missing items.

Respondent type. A statistically significant F test indicated that missingness rates in the Employment and

Table 5. Mean Percent Leave-Behind Questionnaire Item-Level Missingness (SD) by Age–Gender Categories

	62–69 years	70–79 years	80–90 years
Males	2.81 (6.47)	4.91 (11.03)	8.27 (13.65)
Females	3.17 (6.12)	6.07 (11.55)	8.33 (13.62)

Finances domain were associated with Wave 1 nonresponse. As shown in Table 4, partners and Wave 1 nonrespondents had higher rates of item-level missingness than Wave 1 respondents, and the former two groups did not differ significantly, p (Scheffé adjustment) = .644.

Biomeasure path. A nonsignificant F test indicated no biomeasure path differences in rates of missingness in the Employment and Finances domain.

Prevalence in the LBQ.—Responses were missing for 4.77 items on average ($SD = 9.58$). Table 5 displays missingness rates by age–gender categories and shows that missingness ranged from 2.81% to 8.33% in these groups. Among those who returned the LBQ, the majority of respondents (88.6%) were missing 10 or fewer items, and 30.4% completed the LBQ entirely. The missingness rate of 4.97% ($SD = 9.97$), although higher than in the CAPI, is still not as high as that reported by Eekhout and coworkers (2012) which, as was noted earlier, averaged 11%. Moreover, in NSHAP, this was the rate of item-level missingness in a leave-behind mail-back questionnaire, not an in-person interview of the type reviewed by Eekhout and coworkers (2012), indicating that the leave-behind mode of administration can elicit relatively complete responses among those who attempt to finish it. Nevertheless, sacrifices are made with this mode of administration: Item-level missingness in the LBQ exceeded 70% in a small number of respondents, whereas the highest rate of missingness in the CAPI approached only 50%. Table 6 provides the mean proportion of missing items within domains of the LBQ across the 96 items included in the analyses.

Among the domains of LBQ variables, the highest rate of missing items was in the Fertility domain ($M = 9.8\%$, $SD = 25.7$),

Table 6. Mean Percent of Leave-Behind Questionnaire (LBQ) Items Missing in the LBQ Attitudes domain (10 items) and Negative Binomial Regression Coefficients Estimating the Log Number of Items Missing by Predictor Category

Predictor	% of items missing (linearized SE)	B (linearized SE)	p
Age group*			
62–69 years (ref)	5.58 (0.57)	—	
70–79 years	8.67 (0.53)	0.44 (0.12)	<.001
80–90 years	13.39 (0.61)	0.88 (0.11)	<.001
Gender*			
Male (ref)	5.93 (0.52)	—	
Female	10.38 (0.41)	0.56 (0.09)	<.001
Race/Ethnicity			
White/Caucasian (ref)	8.02 (0.43)	—	
Black/African American	10.30 (1.00)	0.25 (0.11)	.033
American Indian	11.66 (2.87)	0.37 (0.25)	.143
Asian or Pacific Islander	7.03 (2.38)	–0.13 (0.35)	.711
Other	8.21 (1.13)	0.02 (0.14)	.867
Education*			
Less than high school (ref)	10.96 (0.97)	—	
High school/Equivalent	8.72 (0.52)	–0.23 (0.10)	.023
Vocational certification	8.17 (0.51)	–0.29 (0.10)	.004
Bachelor's degree or more	6.27 (0.49)	–0.56 (0.11)	<.001
Marital status*			
Married (ref)	5.37 (0.33)	—	
Living with a partner	4.98 (1.53)	–0.08 (0.31)	.806
Separated	8.03 (2.56)	0.40 (0.31)	.197
Divorced	11.98 (1.08)	0.80 (0.08)	<.001
Widowed	14.77 (0.56)	1.01 (0.07)	<.001
Never married	13.82 (1.20)	0.94 (0.11)	<.001
Living arrangements*			
Partnered, living alone (ref)	4.56 (1.44)	—	
Partnered, living with someone	5.29 (0.33)	0.15 (0.32)	.641
Unpartnered, living alone	15.34 (0.59)	1.21 (0.31)	<.001
Unpartnered, living with someone	14.32 (0.89)	1.15 (0.34)	.001
Self-rated physical health			
Poor	7.64 (1.83)	—	
Fair	9.53 (0.70)	0.22 (0.24)	.362
Good	8.61 (0.52)	0.12 (0.24)	.618
Very good	7.66 (0.51)	0.002 (0.25)	.993
Excellent	7.09 (0.77)	–0.07 (0.21)	.727
Self-rated mental health			
Poor	9.34 (1.51)	—	
Fair	8.81 (0.89)	–0.06 (0.20)	.767
Good	9.12 (0.70)	–0.02 (0.19)	.901
Very good	8.03 (0.50)	–0.15 (0.18)	.400
Excellent	7.07 (0.62)	–0.28 (0.21)	.182
Cognitive functioning*			
Continuous measure (0–20)		–0.06 (0.01)	<.001
Respondent type*			
Wave 1 respondent (ref)	8.87 (0.39)	—	
Wave 1 nonrespondent	9.25 (1.79)	0.04 (0.19)	.823
Partner	5.52 (0.52)	–0.47 (0.10)	<.001
Biomeasure path			
1	8.99 (0.67)	—	
2	7.70 (0.55)	–0.16 (0.10)	.118
3	7.31 (0.57)	–0.21 (0.12)	.083
4	8.30 (0.80)	–0.08 (0.11)	.489
5	9.03 (1.09)	0.004 (0.13)	.973
6	8.14 (0.74)	–0.10 (0.10)	.349

Note. *Global F test is significant, $p < .05$.

followed by the Attitudes domain ($M = 8.3\%$, $SD = 13.90$), the Childhood background domain ($M = 6.4\%$, $SD = 14.5$), and “Personality” in the Thoughts and Feelings domain ($M = 5.9\%$, $SD = 19.6$). All other domains were missing data on fewer than 5% of the items (range = 1.2% for the Social Relationships and Activities domain to 4.9% for the Health domain).

Predictors of item-level missingness: LBQ.—For example purposes, we chose to model missingness in the Attitudes domain because it has one of the higher rates of missingness, and the items in this domain constitute one of the primary research foci in NSHAP. Results are displayed in Table 6.

Sociodemographic factors. Negative binomial regression models revealed statistically significant F tests indicating that missingness in the Attitudes domain was associated with age, gender, education, marital status, and living arrangements, but not with race–ethnicity. Regression coefficients are shown in Table 6, and patterns of effects are summarized here. For *age*, item-level missingness in the Attitudes domain increased with each increment in age category. Item-level missingness was higher among women than men, as was also observed for item-level missingness in the CAPI Employment and Finances domain. For *education*, item-level missingness decreased with each increment in education category. For *marital status*, the predominant effect was a lower rate of missingness in married than in divorced, widowed, and never married respondents. The latter three groups were also revealed in post hoc comparisons to have higher rates of missingness than those living with a partner, ps (Scheffé adjustments) $< .003$. For *living arrangements*, the predominant difference was between partnered and unpartnered respondents, regardless of living arrangements (see Table 4). Post hoc comparisons confirmed that item-level missingness was greater in unpartnered than partnered respondents whether they lived alone or with someone, ps (Scheffé adjustments) $< .001$.

Health status and cognitive functioning. Negative binomial regression models revealed a statistically significant F test indicating that missingness in the Attitudes domain was associated with cognitive functioning, $p < .001$, but not with self-rated physical or mental health, $ps > .06$. *Cognitive functioning* showed an inverse association with item-level missingness; as shown in Table 4, each one-unit increase in cognitive functioning was associated with a log 0.06 item decrease in missingness.

Respondent type. A statistically significant F test indicated that the missingness rate in the Attitudes domain was associated with Wave 1 nonresponse, $p < .001$. As shown in Table 6, this effect was largely attributable to lower rates of item-level missingness in partners than in Wave 1 respondents. The comparison between partners

and Wave 1 nonrespondents was not reliable, p (Scheffé adjustment) = .103.

Biomeasure path. Negative binomial regression models revealed a nonsignificant F test, $p = .541$, indicating that missingness in the Attitudes domain was not associated with biomeasure path. Alternatively stated, assignment to biomeasure path did not seem to have a differential effect on item-level missingness in the LBQ.

Treatment.—Rates of item-level missingness less than 5%, a rate observed for many of the variables in the CAPI and LBQ domains assessed here, are sometimes considered “inconsequential” (Roth, 1994). That is, complete case analyses using CAPI or LBQ data from NSHAP Wave 2 may still perform well. However, missing data could be consequential in multivariate analyses if there is no or little overlap in the missing variables, and a complete case analysis is left with only a small proportion of cases from the original sample.

Although there are many *ad hoc* methods of handling missing item-level data (e.g., complete case analysis, single imputation, or substituting the average of available items for the total scale score), multiple imputation is the most commonly used formal approach to deal with item-level missingness. MICE is a flexible approach, as described earlier for unit-level nonresponse. Our analyses revealed significant predictors of item-level missingness in two domains of potential interest to researchers: Employment and Finances, and Attitudes toward Touch and Sexuality. Accordingly, for Employment and Finances, variables that should be considered candidates for inclusion in imputation models include age, gender, race—ethnicity, education, marital status, self-rated physical and mental health, cognitive functioning, and respondent type. For Attitudes toward Touch and Sexuality, variables that should be considered candidates for inclusion in imputation models include age, gender, education, marital status, and living arrangements, cognitive functioning, and respondent type. Missingness on items in other domains may require a different set of variables that control for the observed sources of missingness.

DISCUSSION

Wave 2 of NSHAP exhibits low rates of unit-level and item-level missingness relative to extant epidemiological surveys. Nevertheless, the nonreturn rate of approximately 11% of the LBQs warrants attention to the causes of nonreturn, some of which may be considered unexpected and that point to potentially important additional effects. For instance, contradicting the notion that people whose time is more valuable, at least in economic terms (e.g., higher income), are less likely to respond, LBQ nonresponse in NSHAP in Wave 2 is higher among those with less education, lower income, and at disproportionate disadvantage.

Second, those who are married or have a cohabiting spouse or partner were more likely to return the LBQ and less likely to skip individual items. In the case of the LBQ, this may have been due to an influence or contagion effect among spouses and partners who were both included in the study. On the other hand, being married might be a marker for certain individual characteristics that are also predictive of compliance to the request for the return of the LBQ and to the expectation that all items be completed. Future research on this issue might include information on respondent type and a comparison of those who were randomized to receive a partner interview versus those who were not, as well as a comparison with non-response among those who become widowed between waves.

Third, nonresponse rates were higher among those with poorer cognitive function. This is just one possible effect that poor cognitive function may have; others may include a longer interview time, poorer comprehension, and less accurate responses. This becomes an increasing concern as the NSHAP cohort ages. Beginning in Wave 2, a cognitive measure was introduced that will permit examining these concerns in greater detail.

Fourth, the actigraphy module appears to have had an adverse effect on the LBQ return rate, though it is unclear whether this was due to respondent fatigue or simply interference between the two instruments, both of which were in the field at the same time. Fortunately, this does not contribute a systematic bias because the module was administered to a randomly chosen subset of the sample.

Fifth, those with poorer health, and especially those with poorer mental health, were less likely to respond. This means that, in general, marginal estimates of health measures will have some upward bias, and estimates of the association between health measures and covariates may also have some bias.

Knowing some of the causes of missingness assists analysts in applying an appropriate treatment of missing cases to minimize bias in parameter estimates, to maximize statistical power, and to ensure generalizability to the population of interest. Item-level nonresponse in CAPI and LBQ, although not MCAR, were at sufficiently low rates of missingness in any given domain (or variable) that, for bivariate associations, the results of complete case analysis may not differ much from a more comprehensive approach such as multiple imputation. Although there is no universal definition of what is a sufficiently “low” rate of missingness when complete case analysis may be adequate, some have suggested that even when there is only 5% missing data, there are advantages to using multiple imputation (Schafer & Graham, 2002). Thus, in multivariate analyses where the proportion of missingness is increased due to nonoverlapping missingness patterns in the variables, missingness needs to be addressed more formally with methods such as multiple imputation, IPW, or a FIML-based approach.

When some component items of a scale (e.g., depression) are missing, it is common practice to estimate the score by

averaging the available items. Graham (2009) found that a multi-item summary scale score can be acceptably based on partial data if data are available for at least half of the variables and the variables have a relatively high coefficient alpha. However, Schafer and Graham (2002) argue that this approach is not justified theoretically from the sampling or likelihood perspective and may produce biased results, especially if the reliability of the scale is low (e.g., $\alpha < .70$). They recommend using multiple imputation for the individual missing items which can then be combined to calculate the overall score. Such an item-by-item approach has the advantage of using all available information from the non-missing items which may be highly correlated with the missing items. Notably, because MI preserves the intercorrelations between the items, non-missing items on the scale are good choices to include in multiple imputations for individual missing items even if items are poorly correlated (Schafer & Graham, 2002).

Some summary scale scores will persist in missingness, however, and multiple imputation can then be used to impute at the scale level using other information related to missingness. The creation of summary scores can also minimize the number of variables needed in the imputation model, which Graham (2009) argues should be kept less than 100. It should be noted that IRT or full information likelihood methods can accommodate missing items within a scale, and are therefore a better alternative than casewise deletion. Of course, under MNAR, they will still be susceptible to nonresponse bias.

Variable selection for imputation models has been amply discussed in prior literature, and recommended practices consistently involve inclusion of the covariates and outcome from the analysis model (including any interactions), variables involved in the survey design, and predictors of the missing/incomplete variable. Including predictors of the missing variable increases the plausibility of the MAR assumption and consequently reduces bias in estimates (White, Royston, & Wood, 2011). Including survey design features (e.g., weights, strata, and clustering) in imputation models also reduces estimation bias (Reiter, Raghunathan, & Kinney, 2006) and ensures that valid inferences can be drawn from analyses of the multiply imputed data (Schenker et al., 2006). Following the recommendation for NSHAP data articulated by O’Muircheartaigh and coworkers (2009, 2014), we advise that design features be taken into account in multiply imputed data models.

Missingness and Wave 1 Data

An additional source of missing data in Wave 2 may arise when analysts attempt to examine changes in outcomes from Wave 1 to Wave 2. Specifically, analyses of changes in outcomes will be limited to those measures that were obtained in both Waves. In Wave 1, some biomeasures were collected only among a randomized subset of respondents

(i.e., blood spots, Orasure DNA assay, measures of physical and sensory function), meaning that change in these measures in Wave 2 will be restricted to a subset of respondents. Because missingness in this instance is deliberate and random, one can safely ignore it and focus attention on item- or measure-level missingness within those respondents who were administered particular modules of the protocol.

CONCLUSION AND RECOMMENDATIONS

We recommend that analyses carried out using NSHAP data implement a method designed to handle missing data at the unit and item levels. Multiple imputation (MICE in particular) and IPW methods are probably easiest for the average researcher, and numerous guides exist to help researchers through the process. FIML methods are also appropriate (Larsen, 2011) but are generally implemented in a structural equation modeling framework and may therefore have limited utility in analytic contexts that are not easily formulated as SEMs (Graham, 2009). Importantly, we recommend to authors and reviewers that unit- and/or item-level missingness be documented in publications based on NSHAP data, that complete case analysis be avoided unless well justified, and that methods of handling missingness are fully explained.

KEY POINTS

- Wave 2 of NSHAP exhibits low rates of unit-level and item-level missingness relative to extant epidemiological surveys.
- Neither unit-level nor item-level missingness in NSHAP is MCAR, indicating that complete case analyses may be inappropriate.
- Variables that explain missingness include not only characteristics of the sample (e.g., respondent type) but also features of the study design (e.g., biomeasure path to which respondents were assigned; strata; clustering).
- Sample weights provided with NSHAP data do not adjust for nonreturn of the LBQ, and analyses involving the LBQ therefore require special treatment.
- Item-level missingness, although low overall, is likely to be an issue in multivariate analyses with nonoverlapping patterns of missingness.
- Multiple imputation by chained equations (MICE) and inverse probability weighting (IPW) are appropriate and flexible methods readily available in many standard statistical software packages and are recommended to handle unit- and item-level missingness. FIML methods are also appropriate but may be limited to SEM techniques.

FUNDING

This work was supported by the National Institutes of Health including the National Institute on Aging, the Division of Behavioral and Social Sciences Research for the National Health, Social Life, and Aging

Project (NSHAP) (grant numbers R01 AG021487, R37 AG030481), and the NSHAP Wave 2 Partner Project (R01 AG033903) and by NORC, which was responsible for the data collection. Coauthor, J. Wong was supported by a National Institute on Aging Predoctoral Traineeship (T32 AG00243).

ACKNOWLEDGMENTS

The authors thank Philip Schumm for thoughtful feedback on an earlier draft of this manuscript.

CORRESPONDENCE

Correspondence should be addressed to Louise C. Hawkey, PhD, Academic Research Centers, NORC at the University of Chicago, 1155 E 60th Street, Chicago, IL 60637. E-mail: Hawkey-louise@norc.org.

REFERENCES

- Allison, P. D. (2001). *Missing data*. Thousand Oaks, CA: Sage.
- Carpenter, J. R., Kenward, M. G., & Vansteelandt, S. (2006). A comparison of multiple imputation and doubly robust estimation for analyses with missing data. *Journal of the Royal Statistical Society, Series A*, *169*, 571–584. Retrieved from <http://dx.doi.org/10.1111/j.1467-985x.2006.00407.x>
- Eekhout, I., De Boer, J. R., Twisk, J. W. R., De Vet, H. C. W., & Heymans, M. W. (2012). Missing data: A systematic review of how they are reported and handled. *Epidemiology*, *23*, 729–732. Retrieved from <http://dx.doi.org/10.1097/ede.0b013e3182576cdb>
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576. Retrieved from <http://dx.doi.org/10.1146/annurev.psych.58.110405.085530>
- 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*.
- Juster, F. T., & Smith, J. P. (1997). Improving the quality of economic data: Lessons from the HRS and AHEAD. *Journal of the American Statistical Association*, *92*, 1268–1278. Retrieved from <http://dx.doi.org/10.1080/01621459.1997.10473648>
- Larsen, R. (2011). Missing data imputation versus full information maximum likelihood with second-level dependencies. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*, 649–662. Retrieved from <http://dx.doi.org/10.1080/10705511.2011.607721>
- Lauderdale, D. S., Schumm, L. P., Kurina, L. M., McClintock, M., Thisted, R. A., Chen, J.-H., & Waite, L. (2014). Assessment of sleep in the National Social Life, Health and Aging Project [under review]. *Journals of Gerontology, Psychological Sciences and Social Sciences*.
- Little, R. J. A. (1988). Missing-data adjustments in large surveys. *Journal of Business & Economic Statistics*, *6*, 287–296. Retrieved from <http://dx.doi.org/10.1080/07350015.1988.10509663>
- Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York, NY: Wiley.
- Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York, NY: Wiley.
- O'Muircheartaigh, C., Eckman, S., & Smith, S. (2009). Statistical design and estimation for the National Social Life, Health and Aging Project (NSHAP). *Journals of Gerontology, Psychological Sciences and Social Sciences*, *64*, i12–i19. Retrieved from <http://dx.doi.org/10.1093/geronb/gbp045>
- O'Muircheartaigh, C., English, N., Pedlow, S., & Kwok, P. K. (2014). Sample design, sample augmentation, and estimation for Wave II of the National Social Life, Health and Aging Project (NSHAP) [under review]. *Journals of Gerontology, Psychological Sciences and Social Sciences*.
- Reiter, J., Raghunathan, T., & Kinney, S. (2006). The importance of modeling the sampling design in multiple imputation for missing data. *Survey Methodology*, *32*, 143–149.
- Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology*, *47*, 537–560. Retrieved from <http://dx.doi.org/10.1111/j.1744-6570.1994.tb01736.x>

- Royston, P., & White, I. R. (2011). Multiple imputation by chained equations (MICE): Implementation in Stata. *Journal of Statistical Software*, *45*, 1–19.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147–177. Retrieved from <http://dx.doi.org/10.1037//1082-989x.7.2.147>
- Schenker, N., Raghunathan, T. E., Chiu, P.-L., Makuc, D. M., Zhang, G., & Cohen, A. J. (2006). Multiple imputation of missing income data in the National Health Interview Survey. *Journal of the American Statistical Association*, *101*, 924–933. Retrieved from <http://dx.doi.org/10.1198/016214505000001375>
- Seaman, S. R., & White, I. R. (2013). Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research*, *22*, 278–295. Retrieved from <http://dx.doi.org/10.1177/0962280210395740>
- Shega, J. W., Sunkara, P. D., Kotwal, A., Kern, D. W., Leitsch, S. L., McClintock, M. K., . . . Dale, W. (2014). Measuring cognition: The Chicago Cognitive Function Measure (CCFM) in the National Social Life, Health and Aging Project (NSHAP), Wave 2 [under review]. *Journals of Gerontology, Psychological Sciences and Social Sciences*.
- Smith, J., Fisher, G., Ryan, L., Clarke, P., House, J., & Weir, D. (2013). *Psychosocial and lifestyle questionnaire, 2006–2010. Documentation Report Core Section LB*. Ann Arbor, MI: Institute for Social Research.
- St. Clair, P., Bugliari, D., Campbell, N., Chien, S., Hayden, O., Hurd, M., . . . & Zissimopoulos, J. (2011). *RAND HRS Data Documentation, Version L*. Santa Monica, CA: RAND Center for the Study of Aging.
- Van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research*, *16*, 219–242. Retrieved from <http://dx.doi.org/10.1177/0962280206074463>
- White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Tutorial in Biostatistics*, *30*, 377–399. Retrieved from <http://dx.doi.org/10.1002/sim.4067>
- Yan, T., & Curtin, R. (2010). The relation between unit nonresponse and item nonresponse: A response continuum perspective. *International Journal of Public Opinion Research*, *22*, 535–551. Retrieved from <http://dx.doi.org/10.1093/ijpor/edq037>