



# HHS Public Access

Author manuscript

*J Marriage Fam.* Author manuscript; available in PMC 2015 June 23.

Published in final edited form as:

*J Marriage Fam.* 2013 February ; 75(1): 221–234. doi:10.1111/j.1741-3737.2012.01021.x.

## Methods for Handling Missing Secondary Respondent Data

Rebekah Young and David R. Johnson\*

Rebekah Young: rlyoung@uw.edu; David R. Johnson: drj10@psu.edu

Department of Biostatistics, University of Washington, Box 354922, Seattle, WA 98115

\*Department of Sociology, The Pennsylvania State University, 211 Oswald Tower, University Park, PA 16802

### Abstract

Secondary respondent data are underutilized because researchers avoid using these data in the presence of substantial missing data. We reviewed, critically evaluated, and tested potential solutions to this problem. Five strategies of dealing with missing partner data are reviewed: complete case analysis, inverse probability weighting, correction with a Heckman selection model, maximum likelihood estimation, and multiple imputation. Two approaches were used to evaluate the performance of these methods. First, we used data from the National Survey of Fertility Barriers (N = 1,666) to estimate a model predicting marital quality based on characteristics of women and their husbands. Second, we conducted a simulation based on these data testing the five methods and compared the results to estimates where the true value was known. We found that the maximum likelihood and multiple imputation methods were advantageous because they allow researchers to utilize all of the available information as well as produce less biased and more efficient estimates.

### Keywords

maximum likelihood; missing data; multiple imputation; nonresponse; secondary respondents

---

Many of the commonly used data sets in family research have secondary respondent questionnaires linked to the primary respondent. For example, the Add Health study includes a parental interview added to the main adolescent sample (Harris, 2008) and the National Survey of Families and Households (NSFH) include surveys of the primary respondent's cohabiting partner, spouse, or other related householder (Bumpass & Sweet, 1987). Combining information from the primary and secondary respondents greatly expands the range of questions researchers can address. The relationship between wives' employment and marital happiness, for instance, is undetected unless joint preferences of spouses are taken into account (Bumpass, 1990).

The utility of available secondary respondent data is limited when data collection efforts failed to obtain interviews with a large proportion of the secondary respondents, resulting in substantial amounts of missing dyadic data. A common response is to only use those cases where interviews with the secondary respondent were obtained or to opt not to use the dyadic data at all. A practical barrier to greater use of secondary respondent data is that few accessible methodological guidelines that deal with this problem are available to family

researchers. An exception is a study on nonresponse of secondary respondents by Kalmijn and Liefbroer (2010) who applied two methods to one example to show that nonresponse minimally biases substantive estimates. This study has limited utility, however, because several commonly used methods were not evaluated.

In this paper we compare five estimation strategies for handling missing secondary respondent data: complete case analysis, inverse probability weighting, maximum likelihood, multiple imputation, and Heckman selection correction. Using an empirical model similar to those found in family research, we compare the consistency of substantive conclusions from regression analyses across the five strategies. If the methods produce similar results, then choice of approach is less critical for valid inferences. Different results would show the findings were not robust to the choice of methods but it would not be possible to assess a preferred method as the true model is not known. To compare the approaches to the true parameters, we also use a simulation to assess each approach. The goal of our paper is to help researchers make informed choices when analyzing dyadic data.

## Background

The subject of missing data has been widely addressed in a growing body of literature that is accessible to family researchers (e.g., Acock, 2005; Acock, 2012; Graham, 2009). Johnson and Young (2011) offered a number of practical recommendations for handling item-level missing values. Missing all data for a secondary respondent, nevertheless, is often treated as being conceptually different than missing responses to a subset of survey items. Missing information due item-level nonresponse and missing all data on a respondent might initially appear to be unrelated issues that must be solved in different ways. Actually, these issues are closely related.

Nonresponse is often separated into two conceptual categories, unit nonresponse and item nonresponse. Unit nonresponse indicates the failure to obtain any information at all from the respondent, whereas item nonresponse indicates the failure to obtain information about a specific question (Groves et al., 2004). When an interview is completed with a primary respondent but not the associated secondary person, the resulting nonresponse falls somewhere between unit and item nonresponse. Although the secondary respondent provided no information, some information about the secondary respondent may be available from the main respondent. For example, primary respondents are routinely asked to report basic characteristics of other household members.

Researchers have been hesitant to employ the methods used to handle item-level missing data for missing secondary respondents. For example, Heard (2007) and Johnson and Johnson (2009) both use multiple imputation for data missing from the primary respondent variables but drop cases without secondary respondent interviews. Missing secondary respondents are also a missing data problem, and one to which familiar methods can be applied.

### Method for Handling Missing Completely at Random (MCAR) Respondents

**Complete case analysis**—A common method for dealing with missing secondary respondent data is to restrict the analysis to cases with complete data on both the primary and secondary respondents (e.g., Cleveland & Gilson, 2004; Knoester & Haynie, 2005; Neiss & Rowe, 2000). Most statistical software packages use casewise deletion by default, making it a convenient method to use.

Complete case analysis yields biased estimates and incorrect standard errors unless the amount of missing data is small and the missingness is distributed completely at random (Acock, 2005; Allison, 2001; Schafer & Graham, 2002). In the secondary respondent case, missing completely at random (MCAR) means the probability an interview was not obtained is totally random and unrelated to anything else, whether measured or not (Allison, 2001; Acock, 2005; Little & Rubin, 2002). Missing secondary respondents are seldom infrequent and are not likely to be MCAR. For example, in the National Survey of Fertility Barriers (Johnson & White, 2009), interviews were not obtained from roughly half of the secondary respondents. The assumption that secondary respondents are missing completely at random is also unlikely. Johnson and Johnson (2009), for instance, found a number of primary respondent characteristics that predicted nonresponse by the secondary respondent in the National Survey of Fertility Barriers. Smock and Manning (1997) and Sanchez, Manning and Smock (1998), showed that missing partner data were not MCAR in the NSFH. Sassler and McNally (2003) showed that incorrect conclusions about the effects of cohabitation would be reached if the missing partner data in the NSFH was treated with complete case analysis.

Complete case analysis is also a limited technique because it does not allow researchers to use of all known information in their analysis. Removal of cases leads to loss of statistical power—inflating standard errors and decreasing the likelihood of detecting significant effects. This is a more serious problem with small samples. In a study of matched adoptive and biological children, for instance, Neiss and Rowe (2000) had their small sample of 392 reduced by 16% due to missing secondary respondent data. Even in large samples, an optimal missing data strategy should make use of as much information as is available to improve estimates. In sum, complete case analysis falls short of being an optimal method for dealing with missing secondary respondent data because it is likely to produce biased estimates when the MCAR assumption is violated and it discards information from incomplete cases.

### Methods for Handling Missing at Random (MAR) Respondents

Unlike complete case analysis, the next methods discussed here assume that the secondary respondents are missing at random (MAR). Missing at random means that the probability that a secondary respondent interview was not obtained is conditional on one or more observed measures (Allison, 2001; Acock, 2005; Little & Rubin, 2002). Under MAR, the relationship between the missingness and the data can be accounted for by other variables in the dataset.

In practice, the MAR assumption generally cannot be verified and its plausibility is frequently questioned. In particular, researchers worry that the missing data are not missing at random (NMAR). Not missing at random means that the probability that an interview was not obtained depends on characteristics of the secondary respondents even after conditioning on the observed data. For example, in a study of marital quality where wives were the primary respondent and husbands were the secondary, husbands with poor relationships with their wife may be less likely to complete the interview. In practice, most social science data is likely to contain some departures from MAR. Whether these departures are serious, and whether violations cause MAR-based methods to perform poorly, are probably important questions to ask (Schafer & Graham, 2002). In many cases, available evidence suggests that violations of the MAR assumption has minimal impact on estimates and standard errors when MAR-based methods are used (Collins, Schafer & Kam, 2001). For many family researchers, MAR is probably the most practical and realistic assumption to routinely make. Deciding to use casewise deletion because of suspicion that the missing information may not be MAR can yield greater bias because casewise deletion requires an assumption of MCAR, which is even more stringent than the MAR assumption.

**Weighting**—Weighting procedures are a commonly used class of methods for treating missing secondary respondent data. Little and Rubin (2002) discussed several methods for estimating and applying weights and considered this approach to be a modification of complete case analysis. All commonly used statistical software packages can accommodate weighting procedures. The term “weighting” refers to techniques that all speak to the question, ‘How much attention should be paid to each case in the dataset?’ In practice, a weight is a value assigned to each case that indicates how much that case will count in statistical analysis. A weight of 2.00, for example, counts the case twice in the statistical analysis. In weighting for missing secondary respondents, the interviews that were obtained are weighted to represent those that were not obtained.

With inverse probability weighting (IPW) (Robins, Rotnitzky & Zhao, 1995; Scharfstein, Rotnitzky & Robins, 1999) each observed value is given a predicted probability of having been observed—usually estimated by logistic regression—and the inverse of these scores are used to weight the observed data to match the distribution of the full sample. For example, in a survey of adolescents’ mothers and their teenage child we might find that daughters had a .75 probability of taking the survey, whereas sons had only a .25 probability of being interviewed. The cases with sons who completed the survey would be assigned a larger weight than those with daughters to account for fewer cases with teenage sons. The complete cases with sons would be weighted by their inverse probability of responding ( $1/.25 = 4.00$ ), whereas the cases with daughters would be assigned smaller weight values ( $1/.75 = 1.3$ ).

The idea behind IPW is straightforward and intuitively attractive, although questions have been raised about biased and inefficient estimation when the model used to estimate the response probability is incomplete or incorrect (Carpenter et al., 2006). The IPW technique has made an appearance in the family literature (e.g., D’Onofrio et al., 2007; Fergusson, Horwood & Ridder, 2005) though its specific implementation is sometimes unclear. Debate

about the pros and cons of IPW compared to multiple imputation has been discussed in Sharfstein et al. (1999) and Carpenter et al. (2006).

**Maximum Likelihood**—Another general approach to handling missing secondary respondent data is maximum likelihood (ML) estimation. In ML, the covariance matrix among the variables in the analysis model is estimated with a procedure that uses data from both complete and incomplete cases. A number of ML algorithms have been implemented in popular statistical software packages and the technical literature on ML estimation is vast (e.g., Dempster, Laird, & Rubin, 1977; Little & Rubin 2002). Enders (2001) provided an accessible nontechnical review of the differences between widely available ML algorithms. Here, we focus on a ML approach referred to as full-information maximum likelihood (FIML) because this approach is common in structural equation software (e.g. AMOS, Mplus, and SEM in Stata version 12).

In the FIML approach, a complete case likelihood function is estimated using both the complete and incomplete records (Arbuckle, 1995; Enders & Bandalos, 2001). A mathematical equation is then used to maximize the likelihood function (i.e., find the most likely values of distribution parameters for a given set of data). FIML allows information gathered from the primary respondent to be used even when the secondary respondent was not interviewed. Like weighting and multiple imputation (discussed below), FIML assumes the data are MAR. A disadvantage of this approach is that it has only been implemented in structural equation software and there are restrictions on the types of analysis model to which it can be applied.

**Multiple imputation**—Multiple imputation (MI) has emerged as a flexible and widely used alternative for dealing with missing values in family research (Johnson & Young, 2011). In MI, each missing value is replaced with a set of plausible imputed values that are assigned using an appropriate model that incorporates uncertainty about the true value (Little & Rubin, 2002; Rubin, 1987). Statistical software implementations of MI are available in SAS, Stata, and SPSS as well as a number of other packages such as R, IVEware and Amelia. All current MI programs assume that the missing data are MAR, although the MI approach has been found to be robust to violations of this assumption (Rubin, 1996; Schafer & Graham, 2002).

Like maximum likelihood, the MI approach has an advantage over complete case analysis and weighting procedures because it retains all cases in the analysis and uses all known information (Rubin, 1996). Additionally, the Bayesian theory that underlies the MI procedure allows it to be useful in making inferences in small samples even when the proportion of missing values is large (Allison, 2001; Little & Rubin, 2002). The advantage of MI relative to ML is that a complete data matrix is created with the imputation approach. This can help facilitate analysis of the same data by multiple researchers and reduce complications when particular variables are used as both independent and dependent variables (Johnson & Young, 2011). It also permits the use of a wider range of analytic procedures. A disadvantage is the need to generate several imputed datasets, a step not required with ML (Johnson & Young, 2011).

Estimates with MI and FIML have been shown to be equivalent when the input data and models are the same and the number of imputed data sets in MI is sufficiently large (Collins, Schafer & Kam, 2001; Graham, 2003). This equivalence should ease the concern of some researchers who are reluctant to use MI because of the fear that it is making up data because no imputed values are used in the FIML approach.

### Method for Handling Not Missing at Random (NMAR) Respondents

Although the MAR assumption is likely to be a reasonable one in many real-world scenarios, NMAR data are sometimes a valid concern. In clinical trials or treatment studies, in particular, participants may be dropping out of a study for reasons closely related to the outcome being measured. For example, couples attending marriage therapy who do not see their relationship improving may decide to pursue alternative strategies and stop attending the therapy program. The statistical and survey methods literature has not clearly delineated the circumstances where secondary respondent data are likely to be NMAR. Researchers concerned about this issue may wish to explore methods that do not assume MAR, or, see if the conclusions reached are sensitive to plausible NMAR models.

To perform adequately, methods that allow NMAR data generally require explicit specification of the distribution of the missingness and a model for the complete data (Schafer & Graham, 2002). These models are often difficult to handle computationally and may be tricky to implement and replicate in datasets used by different groups of researchers. One of the most promising ways to deal with NMAR data is to take steps to recover the data, such as following up with missing respondents (Schafer & Graham, 2002). For researchers working with secondary data, this is an impractical strategy likely to be prohibited by institutional review boards and data license agreements.

Two modeling approaches to NMAR data are the Heckman selection correction and the use of pattern mixture models. Pattern mixture models apply the substantive model to groups that exhibit different missing data patterns; the goal is to detect group differences parameter estimates of the substantive model (Allison, 2002). Pattern mixture models cannot be applied when a secondary respondent assessment is the outcome variable in a regression model because the missing data pattern composed of the cases where the secondary respondent did not complete the interview would not be estimable. Therefore, we do not test this approach here and merely remind readers that it may be applicable in other situations.

**Heckman selection correction**—Data loss for missing secondary respondent interviews may reflect a selection mechanism that can be modeled using other variables included in the data set. A common strategy has been to use a Heckman two-step procedure to estimate the selection model (Heckman, 1976, 1979). The Heckman method assumes an underlying cause of the missing data and deals with it as a specification error problem, or omitted variable bias. Heckman's two-step procedure is specified by fitting a selection equation (usually a probit model) to estimate the expected error and generating a variable (the inverse Mills ratio:  $\lambda$ ) to represent this error (Greene, 1993). The second step is to add this variable to the main regression model as an additional covariate.  $\lambda$  is expected to remove the sample selection bias from the error term, yielding estimates unbiased by the



nonrandom missing data. The cases with missing secondary respondent data are dropped from the second step but are “accounted for” by the presence of lambda as a covariate in the model.

The two-step Heckman correction requires knowledge about the distribution of the residuals in subgroups with different selection rates to successfully adjust for selection (Little, 1985). The use of residuals to determine selection bias adjustments is not robust; the influence of deviations of observed residuals from their distribution under the assumed model can lead to a poor fit (Little & Rubin, 2002). Although widely criticized and empirically demonstrated to perform poorly under particular selection models, the Heckman correction continues to be defended (Grasdal, 2001; Heckman, 2005a, 2005b; Puhani, 2000; Sobel, 2006). This approach has been applied often in family research despite its sensitivity to incorrect or incomplete model specifications (e.g., DeMaris et al., 2003; Fan & Abdel-Ghany, 2004; Kalmijn & Liefbroer, 2010; McGinnis, 2004).

## Method

We first used an empirical example typical of those found in family research to compare how the five methods of handling missing secondary respondent data influenced the substantive results. A “real data” example tests the sensitivity of the estimated parameters, standard errors, and significance tests to the approach used to handle missing secondary respondents. The tradition of using case studies involving “real world” examples to help shape methodological practices has a long history that spans many disciplines (e.g., Austin & Mamdani, 2005; Erickson, 1978; Gelman & Pardoe, 2007; Johnson & Elliott, 1998; Sassler & McNally, 2003).

The empirical regression model used here examines the effects of health on marital quality for married couples. Previous research has found that marital quality is affected by both the person’s own health and by the health of their spouse (Booth & Johnson, 1994; Yorgason et al., 2008). The findings from these studies may be biased because partners were not interviewed and the health of the spouse was reported by the respondent. The model reported here includes reports from both the husband and the wife.

We used both the primary and associated secondary respondent data from the National Survey of Fertility Barriers (NSFB). The NSFB is a nationally representative random digit dialing telephone survey designed to study social and psychological aspects of fertility barriers. It was conducted between 2004 and 2007 and interviewed 4,699 women ages 25–45 and their partners. For our analysis we restricted the sample to married women and their husbands. Although the survey was designed to interview partners, completed interviews on partners were only obtained for 47% of the women with partners. Johnson and Johnson (2009) found that women whose partner did not respond were different from women whose partner did respond, being more likely to be white and have higher fertility intentions and less likely to have children from a previous marriage. Because the husbands not responding to the survey were not MCAR, accounting for the missing secondary respondent data is necessary to reduce bias in estimates using the partner data.

Our outcome variable, marital quality, was measured by a 3-item scale asked of all married primary respondents and their partners. The three items were: (1) “Have you ever thought your relationship might be in trouble?” (*Yes/No*). (2) “Taking all things together, how would you describe your relationship? Would you say that it is very happy, pretty happy, or not too happy?” (3) “Overall, how satisfied are you with your sexual relationship? Would you say very satisfied, pretty satisfied, or not too satisfied?” The three items were standardized and summed to create the scale. We tested separate models for the husband’s and the wife’s marital quality. The alpha for the marital quality scale was 0.64 for the wife’s report, and 0.63 for the husband’s.

The models tested examined the effects of the woman’s and her husband’s self-reported health on their marital quality. Additional health measures asked of both spouses and included in the models were a depression scale based on a subset of items from the Center for Epidemiological Studies of Depression (CES-D) scale and a self-esteem scale. We also included a count of the number of reported chronic illnesses reported by the wife—not available on the husband’s interview. Marital duration was also included in the model. This relationship of the duration to wife’s marital quality was found to be nonlinear and was represented by a quadratic term in and linear when the husband’s marital quality was the outcome. Some variables were included, such as chronic illnesses, were included because they had a weak or nonsignificant effect to test whether any of the methods were more likely than others to incorrectly identify the variable as significant (a “false-positive”).

Additional “auxiliary” variables, not used in the regression model, were used in generating inverse probability weights, to inform the ML and MI models, and to model sample selection in the Heckman procedure. We selected these auxiliary variables based on their ability to predict whether or not an interview was obtained with the husband. Descriptive information the variables used in the regression models and the auxiliary variables are shown in Table 1.

The amount of missing data among model variables for the primary respondents was mostly small (less than 1%). For the auxiliary variables total family income had the most missing values (6.45%). Each of the four model variables from the husband’s interview had slightly more than 50% missing data, almost entirely due to failure to obtain an interview. The large proportion of missing secondary respondents data allowed for a rigorous comparison of the different estimation strategies.

In the empirical example, of course, we do not know true estimates that would have been obtained if the secondary respondents had been fully observed. While it is helpful to know if the results are sensitive to different approaches, our real data example cannot demonstrate which method performs best. To test which method best fits a true model a simulation is needed which allows us to compare the estimates with missing secondary respondents to those that would have been obtained if all had provided interviews. We designed our simulation to yield a true model that maintained the data structure and patterns of missing data found in the NSFB empirical model used here.



For our simulation we began with the same data set and variables used in our empirical models and retained only the 826 cases where both spouses had been interviewed. Estimates on these 826 cases were used to represent our “true” model. We next developed a selection model to introduce nonresponse by secondary respondents into the data set. We fit a logistic regression model to the full data set with all variables measured on the main primary respondent listed in Table 1 as predictors to predicting whether or not a completed interview was obtained for the secondary respondent. This equation was used to estimate the probability that a secondary respondent with this set of characteristics would have responded to the survey. This response propensity was calculated for all cases, including those where the husband had responded to the survey. This response propensity was then used to assign approximately 50% of the 826 cases to be missing. For example, for cases with a .75 response propensity, 25% were randomly chosen to be “missing” secondary respondents. We generated 1,000 simulated data sets each with different random draws on these propensities. Among the cases used in the simulation, the predicted response propensity had a mean of 55% and ranged between 13% and 99%. The advantage of this approach is that the simulated data has a complex covariance structure and nonresponse selection process that closely mirrors the empirically observed data. This simulation yielded missing data that were largely MAR but contained some departures best described as NMAR. Note that we could only verify these assumptions about the missingness because we assigned the missing data ourselves and the true non-missing values were known; the MAR assumption cannot generally be verified in practice. Allowing the MAR assumption to be violated in a realistic way provided an important test of how the missing data approaches used here might perform in circumstances where NMAR is suspected.

We used seven measures to compare the performance of the missing data methods applied to the simulated datasets. For each independent variable we calculated the difference from the true b-coefficient (i.e., the mean bias) averaged across the 1,000 simulated datasets. Second, we included the standard deviation (SD) of this difference in the 1,000 data sets. Third, we calculated the root mean square error (RMSE), an indicator commonly used in simulation studies to estimate overall bias. The RMSE was calculated as follows:

$$RMSE = \sqrt{\frac{\sum (b_t - b_i)^2}{N}}$$

where  $b_i$  is the estimate (predicted value, or y-hat) for any simulated data set  $i$ ,  $b_t$  is the population (true, or  $y$ ) estimate, and  $N$  is the number of simulated data sets. The fourth measure is the proportion (P) of the 1,000 simulated datasets where the significance of the independent variable in the simulated dataset matched the significance in the true model at the  $p < .05$  level. The measures for each variable in the models are found in Appendix B of the online version of this article on Wiley Interscience.

We also created three summary measures that provide an overall assessment of the performance of each approach to handling the missing secondary respondents. Our first measure averaged across all variables in the model the absolute value of the difference between the predicted standardized coefficient and the true standardized coefficient,

reported as the standardized average coefficient bias. Standardization was an important step to eliminate the different metrics of the independent variables from carrying different weight in the summary measure. The standardized average RMSE was calculated by substituting the standardized coefficients for the b-coefficients in the previous calculation and averaged across all model variables. Finally we averaged across all independent variables the proportion of cases in which there was a match of the significance levels of the simulated and true models.

## Results

### Observed Data

The five missing data approaches were each applied to two ordinary least squares regression models: one that used the wife's report of her marital quality as the outcome and one that used the husband's report of his marital quality as the outcome. Complete case analysis—the first approach—used only the cases that had complete data from both partners. The second approach, inverse probability weighting (IPW), also used only the complete cases but assign weights to each case to adjust for the propensity to be included in the sample. All the model and auxiliary variables shown in Table 1 were used in a logistic regression to estimate the response probability. The third approach was full information maximum likelihood (FIML); all observed information on the variables in the regression plus the auxiliary variables were used to estimate the model with the sem procedure in Stata. The fourth approach used a dataset where all model variables were imputed using sequential chained regression implemented in Stata MI. All model and auxiliary variables were used in the imputation model and 40 datasets were generated. We also evaluated using 100 imputed datasets but based on the fraction of missing information (see Schafer & Olsen, 1998), 40 datasets appeared to be sufficient. The final approach we tested was a two-step Heckman selection model where all variables from the primary respondent and the auxiliary variables listed in Table 1 were used in the selection equation. The results of these analyses are presented in Tables 2 and 3. The coefficients for the selection step (probit model) of the Heckman models are available in Appendix A of the online version of this article on Wiley Interscience.

The results for each approach applied to the model predicting the primary respondent's (wife's) marital quality are presented in Table 2. Sample sizes varied among the approaches. The IPW sample,  $n = 768$ , was the smallest. A limitation of IPW is that cases missing values on any of the variables used to calculate the weights (i.e., the primary respondent and auxiliary variables) must also be excluded from the analysis. With the Heckman procedure, the missing partners are explicitly modeled so missing secondary respondents are not excluded. Missing values on other variables are still excluded from the procedure. The resulting sample size of  $n = 1,523$  still falls short of the goal of using all known information. All cases were used with the FIML and MI procedures ( $n = 1,666$ ).

The pattern of findings was largely similar across approaches. Enough differences occurred, however, to raise concerns about the sensitivity of the results to the approaches used. The standard errors of the variables measured on the main respondent produced by FIML and MI were always smaller relative to those estimated by the other approaches, reflecting the larger

sample sizes used and low amount of missing data for variables from the wife's survey. The Heckman approach raised concerns because it was the only method that did not find the coefficient for the primary respondent's self-rated health to be significant ( $p < .05$ ) and the b coefficient was substantially smaller than found in the other approaches. As the research model was framed to study the relationship between health and marital quality, this difference is substantial as very different conclusions about the effect of health on marital quality would have been reached if only the Heckman correction had been used. The effect of self-esteem also appeared sensitive to the approach used as the Heckman approach found virtually no effect on the outcome, but self-esteem was significant in three other approaches and nearly so in the fourth (IPW).

Table 3 presents results for each approach applied to a model predicting the secondary respondent's (husband's) marital quality. In this model, husband's own health rating significantly ( $p < .05$ ) affected his marital quality in all but the IPW models, and wife's health was not significant in any of the models. All approaches yielded similar results with the exception of the lack of a significant effect of husband's health rating in the IPW model. This coefficient was smaller in magnitude than coefficients obtained by any other approach, but IPW also had the largest standard errors on all variables as weights introduce design effects into the data that increase standard errors (Winship & Radbill 1994).

Overall, our empirical models found that the use of different strategies for dealing with the missing secondary respondent data yielded similar, but not identical, results. Because we do not know the true estimates that would have been obtained if all husbands had responded, it is impossible to judge which of the approaches was the most accurate. We next examine our simulation results where comparison with the true model is possible.

### Simulated Data

The summary results presented in Table 4 show the performance of the five missing data approaches for the models with the primary respondent's marital quality and the secondary respondent's marital quality as the outcome. Detailed results for each independent variable showing the non-standardized bias, standard deviation of the bias, root mean square error, and proportion of p-values correct at the .05 level are available in Appendix B of the online version of the article on Wiley Interscience.

The measure of coefficient bias shows the absolute value of how different the standardized b-coefficients were, averaged across 1,000 simulations, from the true b-coefficient. Larger bias indicates greater average deviation, either over- or under-estimating the true effect size (recall that the absolute value was used so direction of the bias is not shown in the summary measure). If the estimated and true coefficients were the same, the bias would be 0.000.

Overall, the approaches closely matched the true b-coefficient. For the model estimating the primary respondent's marital quality, the Heckman selection model (.020) had the largest average bias. The average bias (.006) was considerably smaller for the other four approaches. Bias was generally greater in all approaches in the when the secondary respondent's marital quality was the outcome when compared to the primary respondent model. This greater levels of bias likely reflects the smaller number of interviews that were

completed with the secondary respondent ( $N = 400$ , averaged across 1,000 datasets) as even when all information was used, there was roughly 50% less known information for the outcome when the secondary compared to the respondent's marital quality was estimated. The IPW model showed the smallest bias (.007), while the Heckman correction and FIML had the largest (both .015).

We next examined the performance of the approaches by comparing the standardized root mean square error (RMSE) for each of the estimates. The RMSE combined the measure of bias and the standard deviation of the bias into one measure of error. A smaller RMSE indicates lower bias and greater consistency of a method; larger RMSEs indicate that a method was less consistent across the 1,000 simulations. This combined measure is an important indicator because a method with a high RMSE could have a low average bias, but be yield highly inconsistent estimates in any given sample.

Comparison of the RMSEs showed that FIML and MI were the least biased and most reliable methods. In the complete case analysis model predicting the primary respondent's marital quality, the overall bias was .006, which was the same average bias as the FIML and MI models. The complete case analysis RMSE (.042), however, was over twice as large as the RMSE from the FIML and MI models (.018). This means that, over 1,000 trials, complete case analysis produced biased estimates of the effect sizes substantially more often than FIML or MI approaches. The Heckman correction estimates had the largest RMSE (.055) of the five methods. The differences in the RMSEs were less pronounced in the secondary respondent's marital quality models, but FIML and MI were still found to perform the best.

Our final strategy for comparing the methods was based on a measure of practical concern -- how often were the coefficients correctly identified as statistically significant ( $p < .05$ ) in the 1,000 simulated data sets? The last column in Table 4 shows the percentage of the time each missing data method got it right, averaged across all coefficients. The closer this measure was to 100%, the better the approach replicated the significance pattern found in the true model.

When the primary respondent's marital quality was the dependent variable, FIML and MI provided the most accurate statistical significance testing of the five methods; both methods matched the true model 90.5% of the time. Complete case analysis had correct p-values 79.3% of the time. The Heckman selection and IPW models identified significance correctly 78.2% and 78.5% of the time, respectively. When the secondary respondent's marital quality was the dependent variable, the significance testing was less accurate overall. Again, this difference is due to the smaller number of known cases on the dependent variable and reflects a substantial reduction in statistical power when interviews are not obtained which affects all the approaches. Here, FIML performed the best and identified significance correctly 69.6% of the time.

## Discussion

Our comparison of five common strategies used to handle missing secondary respondent confirmed that the choice of missing data method can substantively impact findings in family research. In our empirical example, the findings in the complete case analysis were similar to those from other approaches. Because this approach appears to be the most commonly used strategy in family research, it is comforting to know that the conclusions from these studies might not always be biased. This result, however, should not lead researchers to decide that complete case analysis is the best approach. In our models the kinds of husbands who chose to respond to the NSFB did not differ significantly from all husbands for variables used in the regression analysis although selection into the study was significantly related to variables not included in the regression. Wife's self-esteem was the only variable in our substantive model that significantly predicted her husband's willingness to be interviewed. In this example, complete case analysis appeared to be quite comparable to other methods used to account for the missing husbands. This will not always be the case for missing secondary respondents, as illustrated in our simulations. The methodological literature also contains sufficient examples of failures to account for selective response that then yielded biased results when complete case analysis was used (e.g., Allison, 2002). When faced with missing secondary respondents, it would be prudent for family researchers to explore other options.

Comparison of the findings from both the substantive model and the simulation leads us to prefer FIML or MI to deal with missing secondary respondent data. These methods had, on average, the lowest bias and showed the least variability across simulated data sets. They also appeared to be acceptable methods for statistical significance tests. Estimates from the IPW method were generally unbiased, but the larger RMSEs were typically too high to make this the most efficient strategy. Use of weights introduces a design effect, similar to clustering, that increases standard errors and reduces statistical power. We did note in our simulations that IPW seemed to be just as efficient as FIML or MI when the dependent variable in the analysis came from the secondary respondent. This may or may not be the case with real-world data. Our simulation did not create additional item nonresponse on other variables; half of partners were assigned to be missing but the partners who did respond answered nearly all individual questions (>99.9% across the simulations). Adding additional item non-response would have overly complicated our simulation and deviated from the type of missingness observed in the NSFB. If non-trivial levels of item nonresponse from either respondent were present in the data, the IPW model would have probably been even less efficient than our simulation results suggest.

The Heckman selection models were consistently outperformed by complete case analysis, IPW, FIML, and MI. Perhaps it is surprising that the Heckman model fared so poorly, considering it is a method specifically designed for treating NMAR data. In the simulation, we purposely assigned the missing secondary respondents to be a combination of MAR and NMAR data, so the degree of NMAR was realistic. Another reason the Heckman model should have performed well is because, since we assigned the missingness, we know that the selection model was correctly specified. *A priori*, our simulation strategy created data that should have been biased to favor the Heckman method's performance. The other approaches

assume MAR even though there missing mechanism included NMAR data, and the Heckman's selection equation could rarely be specified as exactly as the one used here. A careful reading of the literature, however, shows that our findings are consistent with recent research. Tauchmann (2010), in an article specifically addressing the Heckman model's commonly observed problem with consistency, found that FIML was clearly the more efficient estimation technique and suggested that Heckman's two-stage model might be most useful in situations where FIML is too computationally demanding. Rubin and Little (2002) had also pointed out that the Heckman model relies on untestable assumptions about the error distribution and was very misleading unless these assumptions were correct; they recommended against its use. Finally, Bushway, Johnson & Slocum (2007) argued that the Heckman approach was really only appropriate for very particular and known types of selection. Given the extensive limitations of the Heckman method, and how poorly it performed in our simulations, we recommend that FIML or MI be used even when researchers are concerned about violations of the MAR assumption.

Our study has important implications for how a family researcher might proceed when analyzing survey data with variables reported by a secondary respondent. Even if only a small proportion of the secondary respondents are missing, we recommend that alternatives to complete case analysis be explored. Conducting sensitivity analysis by testing multiple approaches is an excellent option because it provides the ability to assess the sensitivity of the findings to different treatment of the missing data. If the findings are substantively similar, for example, whether complete case analysis or MI is used, then the researcher can be confident that the results are robust. If different missing data methods impact the substantive interpretation, researchers should be cautious in relying on any method. In particular, it is always worth considering the impact of influential data points, non-normal distributions, measurement error on the independent variables, and specification using different independent and auxiliary variables.

Applying several methods may seem unduly daunting at first, but with statistical packages such as Stata, carrying out these methods is very straightforward. The Stata code we used for each of these methods is available in Appendix C of the online version of the article on Wiley Interscience. Researchers will need to do some exploratory analysis of possible variables that may be related to nonresponse by the secondary respondents as these would be needed for all methods other than complete case analysis. For the FIML and MI models, all independent and auxiliary variables can be included to inform the analysis model.

Researchers could also use hybrid versions of the methods presented here as another way to evaluate the sensitivity of the model to the method. All models we tested, with the exception of MI and FIML, only allow the use of cases with no missing values for any of the variables used (not just in the substantive model, but also in the weight construction or selection equation). As variables are added to an IPW model, for example, the sample size could grow increasingly smaller due to item nonresponse. A hybrid approach might involve using MI to impute item-level nonresponse for both primary and secondary respondents and resetting the unit-level nonresponse for secondary respondents back to missing. This would increase the sample sizes for IPW methods, making the standard errors more efficient. Existing research using hybrid approaches looks promising (e.g., Johnson and Johnson, 2009; Qu and



Lipkovich, 2009; Seaman et al., 2012). Further methodological research is needed to evaluate how well these hybrid approaches compare to the more common methods reported here.

Special methods have been developed for dyadic analysis (Kenny et al., 2006) and the issue may be raised as the applicability of our findings to these methods. These methods are designed to handle the correlated error that is likely to exist among two individuals sampled from the same unit. When the dyadic members are distinguishable (as in our example) the usual analysis approach is similar to what we used here except that the two equations are estimated simultaneously with a structural equation or seemingly unrelated regression (sur) model to account for correlated errors. All the approaches we used, with the possible exception of the Heckman correction, should work with these types of models. We did not use a simultaneous solution of the equations in our analysis primarily because of current limitations in estimating sur models with multiply imputed data. Applications of the dyadic analysis models have mostly used complete cases analysis but acknowledge the weakness of the approach (e.g. Cook, 2008; Feerer and Widaman, 2008). Others have used the FIML approach when fitting structural equations (e.g. Kindermann, 2008). Therefore, we see no reason why our findings here would not apply to correlated error models.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## References

- Acock AC. Working with missing values. *Journal of Marriage and Family*. 2005; 67:1012–1028.
- Acock, AC. What to do about missing values. In: Cooper, H., Editor-in-Chief, editor. *APA Handbook of Research Methods in Psychology: Vol. 3. Data Analysis and Research Publication*. 2012.
- Allison, PD. *Missing data*. Thousand Oaks, CA: Sage; 2001.
- Arbuckle, JL. *Amos user's guide [Computer software]*. Chicago: Smallwaters; 1995.
- Austin PC, Mamdani MM. A comparison of propensity score methods: A case-study estimating the effectiveness of post-AMI statin use. *Statistics in Medicine*. 2005; 25:2084–2106. [PubMed: 16220490]
- Booth A, Johnson DR. Declining health and marital quality. *Journal of Marriage and the Family*. 1994; 56:218–223.
- Bumpass LL. What's happening to the family? Interactions between demographic and institutional change. *Demography*. 1990; 26:243–498.
- Bumpass, LL.; Sweet, JA. *National Survey of Families and Households, 1987–1988*. Madison, WI: University of Wisconsin, Center for Demography and Ecology; 1987.
- Bushway S, Johnson BD, Slocum LA. Is the magic still there? The use of the Heckman two-step correction for selection bias in criminology. *Journal of Quantitative Criminology*. 2007; 23:151–178.
- Carpenter JR, Kenward MG, Vansteelandt S. A comparison of multiple imputation with doubly robust estimation for analyses with missing data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 2006; 169:571–584.
- Cleveland HH, Gilson M. The effect of neighborhood proportion of single-parent families and mother-adolescent relationships on adolescents' number of sexual partners. *Journal of Youth and Adolescence*. 2004; 33:319–329.
- Collins LM, Schafer JL, Kam CM. A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*. 2001; 6:330–351. [PubMed: 11778676]

- Cook, WL. Application of the social relations model formulas to developmental research. In: Card, NA.; Selig, JP.; Little, TD., editors. *Modeling Dyadic Data in the Developmental and Behavioral Sciences*. New York: Routledge; 2008.
- DeMaris A, Benson ML, Fox GL, Hill T, Van Wyk J. Distal and proximal factors in domestic violence: A test of an integrated model. *Journal of Marriage and Family*. 2003; 65:652–667.
- Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*. 1977; 39:1–38.
- D’Onofrio BM, Turkheimer E, Emery RE, Harden PK, Slutske WS, Heath AC, et al. A genetically informed study of the intergenerational transmission of marital instability. *Journal of Marriage and Family*. 2007; 69:793–809. [PubMed: 20640183]
- Enders CK. A primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling*. 2001; 8:128–141.
- Enders CK, Bandalos DL. The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*. 2001; 8:430–457.
- Erickson RS. Analyzing one-variable three-wave panel data: A comparison of two models. *Political Methodology*. 1978; 5:151–167.
- Fan JX, Abdel-Ghany M. Patterns of spending behavior and the relative position in income distribution: Some empirical evidence. *Journal of Family and Economic Issues*. 2004; 25:1058–10476.
- Ferrer, E.; Widaman, KF. Dynamic factor analysis of dyadic affective processes with intergroup differences. In: Card, NA.; Selig, JP.; Little, TD., editors. *Modeling Dyadic Data in the Developmental and Behavioral Sciences*. New York: Routledge; 2008.
- Fergusson DM, Horwood LJ, Ridder EM. Partner violence and mental health outcomes in a new zealand birth cohort. *Journal of Marriage and Family*. 2005; 67:1103–1119.
- Gelman A, Pardoe I. Average predictive comparisons for models with nonlinearity, interactions, and variance components. *Sociological Methodology*. 2007; 37:23–51.
- Graham JW. Adding missing-data-relevant variables to FIML-based structural equation models. *Structural Equation Modeling*. 2003; 10:80–100.
- Graham JW. Missing data analysis: Making it work in the real world. *Annual Review of Psychology*. 2009; 60:549–576.
- Grasdal A. The performance of sample selection estimators to control for attrition bias. *Health Economics*. 2001; 10:385–398. [PubMed: 11466801]
- Greene, WH. *Econometric analysis*. 2. New York: Macmillan; 1993.
- Groves, RM.; Fowler, FJ.; Couper, MP.; Lepkowski, JM.; Singer, E.; Tourangeau, R. *Survey methodology*. Hoboken: John Wiley and Sons; 2004.
- Harris, KM. *The National Longitudinal Study of Adolescent Health (Add Health), Waves I & II, 1994–1996; Wave III, 2001–2002*. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill; 2008.
- Heard H. Fathers, mothers and family structure: Family trajectories, parent gender, and adolescent schooling. *Journal of Marriage and Family*. 2007; 69:435–450.
- Heckman JJ. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic Social Measurement*. 1976; 5:475–492.
- Heckman JJ. Sample selection bias as a specification error. *Econometrica*. 1979; 47:53–161.
- Heckman JJ. Rejoinder: Response to Sobel. *Sociological Methodology*. 2005a; 32:135–162.
- Heckman JJ. The scientific model of causality. *Sociological Methodology*. 2005b; 35:1–98.
- Johnson DR, Elliott L. Sampling design effects: Do they affect the analyses of data from the national survey of families and households (NSFH)? *Journal of Marriage and the Family*. 1998; 60:993–1001.
- Johnson K, Johnson DR. Partnered decisions? Infertility and help-seeking in U.S. couples. *Family Relations*. 2009; 58:431–444. [PubMed: 20160961]

- Johnson, DR.; White, L. National Survey of Fertility Barriers [Computer File]. University Park, PA: Population Research Institute, the Pennsylvania State University [distributor]; 2009.
- Johnson DR, Young R. Toward best practices in analyzing datasets with missing data: Comparisons and recommendations. *Journal of Marriage and Family*. 2011; 73:926–945.
- Kalmijn M, Liefbroer AC. Nonresponse of Secondary Respondents in Multi-Actor Surveys: Determinants, Consequences, and Possible Remedies. *Journal of Family Issues*. 2010; 32:735–766.
- Kenny, DA.; Kashy, DA.; Cook, WL. *Dyadic Data Analysis*. New York: The Guilford Press; 2006.
- Kindermann, TA. Can we make causal inferences about the influence of children’s naturally existing social networks on their school motivation?. In: Card, NA.; Selig, JP.; Little, TD., editors. *Modeling Dyadic Data in the Developmental and Behavioral Sciences*. New York: Routledge; 2008.
- Knoester C, Haynie DL. Community context, social integration into family, and youth violence. *Journal of Marriage and Family*. 2005; 67:767–780.
- Little RJA. A note about models for selectivity bias. *Econometrica*. 1985; 53:1469–1474.
- Little, RJA.; Rubin, DB. *Statistical analysis with missing data*. Hoboken: John Wiley & Sons, Inc; 2002.
- McGinnis SL. Cohabiting, dating, and perceived costs of marriage: A model of marriage entry. *Journal of Marriage and Family*. 2004; 65:105–116.
- Neiss M, Rowe DC. Parental education and child’s verbal IQ in adoptive and biological families in the National Longitudinal Study of Adolescent Health. *Behavior Genetics*. 2000; 30:487–495. [PubMed: 11523707]
- Puhani PA. The Heckman correction for sample selection and its critique. *Journal of Economic Surveys*. 2000; 14:53–67.
- Qu Y, Lipkovich I. Propensity score estimation with missing values using multiple imputation missingness pattern (MIMP) approach. *Statistics in Medicine*. 2009; 28:1402–1414. [PubMed: 19222021]
- Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*. 1995; 90:106–129.
- Rubin, DB. *Multiple imputation for nonresponse in surveys*. New York: Chichester; 1987.
- Rubin DB. Multiple imputation after 18+ years. *Journal of the American Statistical Association*. 1996; 91:473–489.
- Sanchez L, Manning WD, Smock PJ. Sex-specialized or collaborative mate selection. Union transitions among cohabitators. *Social Science Research*. 1998; 27:280–294.
- Sassler S, McNally J. Cohabiting couples’ economic circumstances and union transitions: A re-examination using multiple imputation techniques. *Social Science Research*. 2003; 32:553–578.
- Schafer JL, Olsen MK. Multiple imputation for multifariate missing-data problems: A data analyst’s perspective. *Multivariate Behavioral Research*. 1998; 33:545–571.
- Schafer JL, Graham JW. Missing data: Our view of the state of the art. *Psychological Methods*. 2002; 7:147–177. [PubMed: 12090408]
- Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for nonignorable drop-out using semi-parametric nonresponse models (with comments). *Journal of the American Statistical Association*. 1999; 94:1096–1146.
- Seaman SR, White IR, Copas AJ, Li L. Combining multiple imputation and inverse-probability weighting. *Biometrics*. 2012; 68:129–137. [PubMed: 22050039]
- Smock PJ, Manning WD. Cohabiting partners’ economic circumstances and marriage. *Demography*. 1997; 34:331–341. [PubMed: 9275243]
- Sobel ME. Discussion: ‘The scientific model of causality’. *Sociological Methodology*. 2006; 35:99–134.
- Tauchmann H. Consistency of Heckman-type two-step estimators for the multivariate sample-selection model. *Applied Economics*. 2010; 42:3895–3902.

Winship C, Radbill L. Sampling weights and Regression Analysis. *Sociological Methods and Research*. 1994; 23:230–257.

Yorgason JA, Booth A, Johnson DR. Health, disability, and marital quality: Is the association different for older persons? *Research on Aging*. 2008; 30:623–648.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 1**  
 Descriptive Information for Analysis Variables using Complete Case Analysis

Variables	Mean	SD	Min	Max	N	% Missing
<i>Used in regression models</i>						
<i>Primary respondent</i>						
Marital quality	0.00	2.29	-6	3	1,640	1.56
Self-rated health	1.18	0.67	0	2	1,663	0.18
Years married	8.57	6.16	0	34	1,666	0.00
Depression	16.38	4.55	10	39	1,665	0.06
Chronic illnesses	0.29	0.56	0	2	1,666	0.00
Self esteem	10.76	1.49	5	12	1,665	0.06
<i>Secondary respondent</i>						
Marital quality	0.00	2.26	-6	3	822	50.66
Self-rated health	1.20	0.66	0	2	825	50.48
Self-esteem	14.95	3.74	10	36	826	50.42
Depression	10.82	1.37	6	12	826	50.42
<i>Used in missing data models</i>						
Years of school	15.15	2.56	11	22	1,664	0.12
Partner years of school	14.83	2.76	11	22	1,651	0.90
White race (non-white=reference)	0.69	0.46	0	1	1,648	1.08
Life satisfaction scale	12.95	2.19	4	16	1,666	0.00
Family income categories	9.21	2.49	1	12	1,558	6.48
Number of children	1.36	1.22	0	4	1,666	0.00
Thought marriage in trouble	0.16	0.37	0	1	1,666	0.00
Year interviewed 2005	0.56	0.50	0	1	1,666	0.00
Year interviewed 2006	0.38	0.48	0	1	1,666	0.00

Note: Descriptive statistics are unweighted. Marital quality is a 3-item standardized scale; complete case alpha = 0.64 for primary respondent and 0.63 for secondary respondent. SD = standard deviation; N = complete case sample size.

**Table 2**  
 Linear Regression Coefficients Predicting Primary Respondents' Marital Quality, Using Six Missing Data Strategies

<i>Independent Variables</i>	Complete Case Analysis ( <i>n</i> = 811)	Inverse Probability Weight ( <i>n</i> = 768)	Heckman Selection ( <i>n</i> = 1,523)	Full Info. Maximum Likelihood ( <i>n</i> = 1,666)	Multiple Imputation ( <i>n</i> = 1,666)
<i>Primary respondent</i>					
Self-rated health	0.359** (0.129)	0.355* (0.154)	0.262 (0.145)	0.345*** (0.091)	0.343*** (0.091)
Years married	0.051 (0.030)	0.053 (0.034)	0.055 (0.034)	0.072*** (0.022)	0.073*** (0.022)
Years married <sup>2</sup>	0.047** (0.014)	0.049** (0.015)	0.050** (0.015)	0.055*** (0.010)	0.055*** (0.010)
Depression	-0.135*** (0.018)	-0.133*** (0.021)	-0.132*** (0.020)	-0.113*** (0.013)	-0.113*** (0.013)
Chronic illnesses	-0.085 (0.144)	-0.203 (0.175)	-0.202 (0.163)	0.065 (0.104)	0.068 (0.103)
Self-esteem	0.125* (0.055)	0.126 (0.064)	0.019 (0.065)	0.085* (0.038)	0.085* (0.038)
<i>Secondary respondent</i>					
Self-rated health	0.390*** (0.115)	0.380** (0.127)	0.334** (0.116)	0.378** (0.123)	0.391*** (0.124)
Depression	-0.022 (0.021)	-0.020 (0.024)	-0.028 (0.021)	-0.029 (0.022)	-0.030 (0.021)
Self-esteem	-0.012 (0.026)	-0.014 (0.065)	-0.030 (0.056)	-0.026 (0.059)	-0.024 (0.057)
Constant	-1.684 (1.243)	-1.854 (1.384)	1.125 (1.463)	-1.780 (1.032)	-1.806 (1.018)
<i>Lambda</i>			-1.800 (0.381)		

Note: Standard errors (SE) in parentheses. Years married<sup>2</sup> is centered and squared. Inverse probability weighted model used robust standard errors. Multiple imputation used *m* = 40 datasets.

\* *p* < .05.

\*\* *p* < .01.

\*\*\* *p* < .001.



**Table 3**  
 Linear Regression Coefficients Predicting Secondary Respondents' Marital Quality, Using Six Missing Data Strategies

<i>Independent Variables</i>	<b>Complete Case Analysis (<i>n</i> = 819)</b>	<b>Inverse Probability Weight (<i>n</i> = 772)</b>	<b>Heckman Selection (<i>n</i> = 1,527)</b>	<b>Full Info. Maximum Likelihood (<i>n</i> = 1,666)</b>	<b>Multiple Imputation (<i>n</i> = 1,666)</b>
<i>Primary respondent</i>					
Self-rated health	0.010 (0.134)	-0.045 (0.146)	-0.015 (0.140)	-0.026 (0.128)	-0.035 (0.135)
Years married	-0.032** (0.012)	-0.031* (0.013)	-0.030* (0.013)	-0.029* (0.013)	-0.029* (0.011)
Depression	-0.055*** (0.019)	-0.071*** (0.022)	-0.058*** (0.019)	-0.046* (0.018)	-0.045* (0.020)
Chronic illnesses	-0.001 (0.150)	0.040 (0.169)	0.000 (0.157)	0.043 (0.143)	0.027 (0.146)
Self-esteem	0.011 (0.057)	0.019 (0.070)	-0.040 (0.063)	0.001 (0.054)	0.001 (0.057)
<i>Secondary respondent</i>					
Self-rated health	0.356** (0.119)	0.262 (0.137)	0.292* (0.121)	0.351** (0.121)	0.349** (0.126)
Depression	-0.150*** (0.021)	-0.160*** (0.025)	-0.154*** (0.022)	-0.153*** (0.022)	-0.154*** (0.022)
Self-esteem	-0.010 (0.058)	-0.028 (0.063)	0.006 (0.059)	-0.016 (0.058)	-0.011 (0.058)
Constant	2.915** (1.019)	3.605*** (1.252)	4.083*** (1.208)	3.038*** (0.998)	2.980*** (1.044)
<i>Lambda</i>			-0.713 (0.372)		

Note: Standard errors (SE) in parentheses. Years married<sup>2</sup> is centered and squared. Inverse probability weighted model used robust standard errors. Multiple imputation used *m* = 40 datasets.

\* *p* < .05.  
 \*\* *p* < .01.  
 \*\*\* *p* < .001.

**Table 4**

## Summary of Simulation Results

	<u>Average standardized</u>		
	<i>Coefficient bias</i>	<i>RMSE</i>	<i>p-values correct at .05 level</i>
<i>Primary respondents' marital quality was dependent variable</i>			
Complete case analysis	.006	.042	79.3%
Inverse probability weight	.006	.045	78.5%
Heckman selection	.020	.055	78.2%
Full info. maximum likelihood	.006	.018	90.5%
Multiple imputation	.006	.018	90.5%
<i>Secondary respondents' marital quality was dependent variable</i>			
Complete case analysis	.012	.036	69.0%
Inverse probability weight	.007	.037	68.7%
Heckman selection	.015	.039	66.6%
Full info. maximum likelihood	.015	.035	69.6%
Multiple imputation	.012	.035	68.8%

*Note:* All results averaged across 1,000 simulations. Average standardized bias = average standardized (Beta) coefficient difference from the true Beta coefficient; average standardized RMSE = average root mean square error of standardized (Beta) coefficients.