Published in final edited form as:
Stat Med. 2010 June 30; 29(14): 1522–1538. doi:10.1002/sim.3902.

Pattern-mixture models for analyzing normal outcome data with proxy respondents

Michelle Shardell[a,*,†], **Gregory E. Hicks**[b], **Ram R. Miller**[a], **Patricia Langenberg**[a], and **Jay Magaziner**[a]

[a]Department of Epidemiology and Preventive Medicine, University of Maryland, Baltimore, MD, U.S.A

[b]Department of Physical Therapy, University of Delaware, 303 McKinley Lab Newark, DE 19716, U.S.A

Abstract

Studies of older adults often involve interview questions regarding subjective constructs such as perceived disability. In some studies, when subjects are unable (e.g. due to cognitive impairment) or unwilling to respond to these questions, proxies (e.g. relatives or other care givers) are recruited to provide responses in place of the subject. Proxies are usually not approached to respond on behalf of subjects who respond for themselves; thus, for each subject, data from only one of the subject or proxy are available. Typically, proxy responses are simply substituted for missing subject responses, and standard complete-data analyses are performed. However, this approach may introduce measurement error and produce biased parameter estimates. In this paper, we propose using pattern-mixture models that relate non-identifiable parameters to identifiable parameters to analyze data with proxy respondents. We posit three interpretable pattern-mixture restrictions to be used with proxy data, and we propose estimation procedures using maximum likelihood and multiple imputation. The methods are applied to a cohort of elderly hip-fracture patients.

Keywords

disability; gerontology; missing data; pattern-mixture models; proxies; sensitivity analysis

1. Introduction

In studies of older adults, researchers aim to identify mutable factors related to disability. Disability is not directly quantifiable, therefore measurement scales have been developed using multiple self-reports, resulting in approximately continuous, normally distributed variables. One disability measure involves summing scales of dependency in performing instrumental activities of daily living (IADLs), e.g. shopping, managing money, etc. [1].

Some study subjects may be unwilling or unable (e.g. due to dementia) to provide self-reports about disability. In this case, a proxy, such as a relative or other caregiver, is asked to respond in the subject's place [2]. In most studies, proxy data are not collected for subjects who provide self-reports. Thus, for each subject, only one of the subject or proxy respondent

*Correspondence to: Michelle Shardell, Department of Epidemiology and Preventive Medicine, University of Maryland, Baltimore, MD, U.S.A. †mshardel@epi.umaryland.edu.

contributes data. Typically, the data are analyzed by singly imputing the missing subject response with the proxy response. This method implicitly assumes that the proxy and subject have equal response distributions for subjects who do not respond [2-4]. At best, when the assumption is valid, this single imputation results in underestimated variances for parameter estimates, because proxy data are treated as perfect correlates of subject data rather than estimates measured with error [4]. When the assumption is false, this approach produces biased parameter estimates [3,4]. Therefore, alternative analytic methods are needed.

Validation studies consisting of subject–proxy pairs have shown that proxies of older adult subjects tend to report worse subject physical disability than subjects themselves [5-8]. However, these assessments can only be generalized to subjects who are able and willing to provide self-reports. The data structure precludes validating proxy responses as surrogates for the subjects who require proxies; i.e. subjects who do not respond [9].

Few statistical methods address data from proxy respondents. Huang *et al.* [10] proposed a multivariate general linear model for cross-over trials that simultaneously models proxy and subject data assuming that subject data are missing at random (MAR) [11]. In aging research, the MAR assumption is implausible, as the sickest and most cognitively impaired subjects are more likely to require proxies than those who are healthy. Snow *et al.* [8] posited a measurement model under the implausible assumption that subject and proxy data are perfectly correlated. Even if true, perfect correlation does not imply unbiased parameter estimation [12]. The challenge of using proxy data to handle missing data is that observed proxy and subject data are not sufficient to identify the distribution of missing subject data.

This paper has two goals. First, by treating the problem as one of missing data, we use pattern-mixture models [13-15] to propose identifying restrictions for the data distribution among subjects for whom only proxy data are available. Second, we use the models to perform estimation using maximum likelihood (ML) or multiple imputation (MI). This approach involves deriving estimates under a given assumption about missing subject data. We also briefly describe how to derive a single estimate by averaging over an analyst-specified distribution of assumptions. Additionally, we extend the methods to use data from proxy–subject validation subsamples, where data from proxies are collected for a random subsample of subjects who provide data. The proposed methods are applied to the second cohort of the Baltimore Hip Studies (BHS-2), a study of physical recovery from hip fracture. Throughout the paper, we focus on studies where subject data are the gold standard rather than studies where proxies and subjects are two raters of a latent construct. Snow *et al.* [8] explicate this distinction.

2. Methods

In this section we introduce methods for studies without validation data.

2.1. Notation and models

Let $Y_{(s)i}$ and $Y_{(p)i}$ denote subject and proxy responses, respectively, of the ith subject–proxy pair, i=1,…,n. Let $R_{(s)i}$ be the binary indicator for the ith subject response, where $R_{(s)i}=1$ when $Y_{(s)i}$ is observed, and $R_{(s)i}=0$ when $Y_{(p)i}$ is observed. Let $Y_{(\mathrm{obs})i}$ be the observed outcome for the ith pair, where $Y_{(\mathrm{obs})i}=Y_{(s)i}R_{(s)i}+Y_{(p)i}(1-R_{(s)i})$; i.e. exactly one of $Y_{(s)i}$ or $Y_{(p)i}$ is observed for the ith pair. Let $X_i=(X_{1i}, \dots, X_{qi})$ be a row vector of q fully observed covariates. The pair $Y_i=(Y_{(s)i},Y_{(p)i})$ is assumed to follow a bivariate normal distribution conditional on X_i with mean vector $(X_i\beta, X_i\alpha)$, where β and α are column vectors of length q, and variance–covariance matrix Σ, where

$$\sum = \begin{pmatrix} \sigma_{(ss)} & \sigma_{(sp)} \\ \sigma_{(sp)} & \sigma_{(pp)} \end{pmatrix},$$

and $\sigma_{(dd')}$=Cov($Y_{(d)i}$,$Y_{(d')i}$ |X_i), for $d,d' \in \{s,p\}$.

We suppress the subscript i in the notation for the time being. The analysis goal is to estimate the regression equation $E[Y_{(s)} | X]=X\beta$. Let $R_{(s)} | X \sim_{\text{ind}}$ Bernoulli ($\pi_{s|x}$), where Prob($R_{(s)}$=1|X)=$\pi_{s|x}$. The distribution of $Y_{(s)}$ can be rewritten as a mixture of those with observed and missing subject responses:

$$f(Y_{(s)}|X)=\pi_{s|x} f(Y_{(s)}|X, R_{(s)}=1)+(1-\pi_{s|x}) f(Y_{(s)}|X, R_{(s)}=0),$$

where $f(Y_{(s)} |X,R_{(s)}=r)$ is normal with mean $X\beta^{(r)}$ and variance, $\sigma^{(r)}_{(ss)}, \sigma^{(r)}_{(dd')}=\text{Cov}(Y_{(d)}, Y_{(d')}|X, R_{(s)}=r)$, for $d,d' \in \{s,p\}$, $r \in \{0,1\}$. In studies with missing data, pattern-mixture modeling would typically proceed by relating $f(Y_{(s)} |X,R_{(s)}=0)$ to $f(Y_{(s)} |X,R_{(s)}=1)$. For example, the assumption that normal $Y_{(s)}$ is MAR is specified via the pattern-mixture restriction $(\beta^{(1)}, \sigma^{(1)}_{(ss)})=(\beta^{(0)}, \sigma^{(0)}_{(ss)})$. Pattern-mixture models have also been proposed to handle nonignorably missing data [13-15]. However, these approaches do not utilize information available from proxy respondents.

Proxy data, when available, are typically used to singly impute the missing subject data. If $f(Y_{(p)} |X,R_{(s)}=r)$ is normal with mean $X\alpha^{(r)}$ and variance $\sigma^{(r)}_{(pp)}$, then imputing the missing subject data with proxy data and using ordinary least squares to regress $Y_{(\text{obs})}$ on X are tantamount to assuming the pattern-mixture restriction

$$(\beta^{(0)}, \sigma^{(0)}_{(ss)})=(\alpha^{(0)}, \sigma^{(0)}_{(pp)}). \tag{1}$$

If equation (1) holds, then single imputation produces unbiased estimates of β and $\alpha_{(ss)}$. Let $\rho^{(r)}_{(sp)}=\sigma^{(r)}_{(sp)}(\sigma^{(r)}_{(ss)}\sigma^{(r)}_{(pp)})^{-1/2}$ denote the correlation of $Y_{(s)}$ with $Y_{(p)}$ conditional on X for those with $R_{(s)}=r$. Unless $\rho^{(0)}_{(sp)}=1$, standard errors of regression parameters will be underestimated using single imputation even if equation (1) holds [4]. If $\beta^{(0)} \neq \alpha^{(0)}$, this approach will produce biased estimates of β.

The benefit of proxy data is that pattern-mixture restrictions relative to MAR need not be specified. In this paper, we develop methods that can use observed subject and proxy data to model $Y_{(s)}$ under a range of assumptions including MAR, equation (1), and departures from MAR and equation (1). We first consider the following assumptions:

$$\sigma^{(1)}_{(ss)}=\sigma^{(0)}_{(ss)}, \tag{2}$$

$$\sigma^{(1)}_{(sp)}=\sigma^{(0)}_{(sp)}, \tag{3}$$

$$\sigma_{(pp)}^{(1)}=\sigma_{(pp)}^{(0)}. \tag{4}$$

When equations (2)-(4) are assumed, $\sigma_{(ss)}^{(0)}$ is identified by $\sigma_{(ss)}^{(1)}$, and $\sigma_{(pp)}^{(1)}$ is identified by $\sigma_{(pp)}^{(0)}$. However, $\sigma_{(sp)}^{(r)}, r \in \{0, 1\}$, cannot be identified by the data and must be specified by the analyst. Equations (2)-(4) imply that $\rho_{(sp)}^{(0)}=\rho_{(sp)}^{(1)}$. Assuming that equations (2)-(4) hold, we posit additional pattern-mixture restrictions to identify $\beta^{(0)}$, and therefore identify β.

2.1.1. Class of selection bias pattern-mixture models—Each model in the class of selection bias pattern-mixture models posits that the mean of $Y_{(s)}$ among subjects with $R_{(s)}=0$ is a location transformation of the mean of $Y_{(s)}$ among subjects with $R_{(s)}=1$. Specifically,

$$\beta^{(0)}-\beta^{(1)}=\lambda_1, \tag{5}$$

where λ_1 is an unidentifiable analyst-specified q-vector that measures the difference in parameters between those with missing and observed $Y_{(s)}$. For normal $f(\cdot|X)$ assuming equation (2), setting $\lambda_1=0_q$, a length-q vector of 0, is equivalent to MAR. Setting $\lambda_1 \neq 0_q$ specifies nonignorable missingness [13-15]. Throughout, we call a model defined by equations (2) and (5) for specified λ_1 Model 1. Model 1 has previously been proposed to handle missing data without proxy data [13], but it is included here to compare and contrast with pattern-mixture models that use proxy data.

2.1.2. Class of proxy bias pattern-mixture models—Each model in the class of proxy bias pattern-mixture models posits that the mean of $Y_{(s)}$ among subjects with $R_{(s)}=0$ is a location transformation of the mean of $Y_{(p)}$ among subjects with $R_{(s)}=0$. Specifically,

$$\beta^{(0)}-\alpha^{(0)}=\lambda_2, \tag{6}$$

where λ_2 is an unidentifiable analyst-specified q-vector that measures degree of bias introduced by proxy data. Setting $\lambda_2=0_q$ simplifies to the assumption of no proxy bias and implies that singly imputing $Y_{(p)}$ for missing $Y_{(s)}$ leads to unbiased estimates of β. However, unless $\sigma_{(pp)}^{(0)}=\sigma_{(ss)}^{(0)}$, single imputation will lead to biased estimates of $\sigma_{(ss)}$. Throughout we denote a model defined by equations (2)-(4) and (6) for specified λ_2 Model 2.

2.1.3. Class of subject-adjusted proxy pattern-mixture models—Each model in the class of subject-adjusted proxy pattern-mixture models posits that the mean of $Y_{(p)}$ does not depend on $R_{(s)}$, conditional on $Y_{(s)}$. Specifically,

$$E[Y_{(p)}|X, R_{(s)}=0, Y_{(s)}]=E[Y_{(p)}|X, R_{(s)}=1, Y_{(s)}]. \tag{7}$$

We note that $E[Y_{(p)}|X, R_{(s)}=r, Y_{(s)}]=X\alpha^{(r)}+\sigma_{(sp)}^{(r)} \times (Y_{(s)}-X\beta^{(r)})/\sigma_{(ss)}^{(r)}, r \in \{0, 1\}$. Define $\gamma^{(r)}=\alpha_{(sp)}^{(r)}/\alpha_{(ss)}^{(r)}, r \in \{0, 1\}$. Equations (2)-(4) imply that $\gamma^{(0)}=\gamma^{(1)}=\gamma$; therefore, equation (7) implies $\alpha^{(0)}-\gamma\beta^{(0)}=\alpha^{(1)}-\gamma\beta^{(1)}=\lambda_3$, an unidentifiable q-vector. As a result,

$$\beta^{(0)}=(\alpha^{(0)}-\lambda_3)/\gamma, \tag{8}$$

a location-scale transformation from $\alpha^{(0)}$. Equation (7) with equations (2)-(4) is analogous to a previously published pattern-mixture restriction for bivariate normal data with only one potentially missing component [14]. However, because both components are never observed at the same time in this case, the restriction in equation (7) is underidentified, hence λ_3. Equation (7) is only useful if $\sigma^{(r)}_{(sp)}\neq 0$. If $\sigma^{(r)}_{(sp)}=0$, then it implies $\alpha^{(1)}=\alpha^{(0)}$, but $\beta^{(0)}$ is left unspecified. We denote a model defined by equations (2)-(4) and (8) for specified λ_3 Model 3.

Model 3 is an extension of the linear non-additive outcome measurement error model with constant variance [16]. When $\lambda_3=0_q$, Model 3 is equivalent to the measurement error model which assumes that $f(Y_{(p)}|X,R_{(s)}=r,Y_{(s)})=f(Y_{(p)}|R_{(s)}=r,Y_{(s)})$. Weaker assumptions such as departures from equation (7) can also be considered. For example, if $\alpha^{(r)}-\gamma\beta^{(r)}=\lambda_3^{(r)}$, then two sets of unidentifiable q vectors $\lambda_3^{(0)}$ and $\lambda_3^{(1)}$ need to be specified. Such a model is more flexible than Model 3, but at the cost of parsimony.

2.2. Estimation: Maximum likelihood

The mean of $Y_{(s)}$ as a function of X is a weighted average of the two missing-data patterns:

$$E[Y_{(s)}|X]=\sum_{r\in\{0,1\}}\pi^r_{s|x}[1-\pi_{s|x}]^{(1-r)}E[Y_{(s)}|X,R_{(s)}=r].$$

Except for low-dimensional categorical X, $\pi_{s|x}$ is usually modeled as a non-linear function of X using, e.g. logistic or probit regression, producing parameters of $E[Y_{(s)}|X]$ that are difficult to interpret. To circumvent this problem, the model is reformulated as mixtures of $f(Y_{(s)},Y_{(p)},X,R_{(s)})=f(Y_{(s)},Y_{(p)}|X,R_{(s)})f(X|R_{(s)})\text{Prob}(R_{(s)}=r)$. X_i is assumed to be multivariate normal to obtain the empirical mean vector and variance–covariance matrix of X_i given $R_{(s)}$, because these quantities are used to estimate linear regression parameters β. Multivariate normal is often not a plausible assumption; however, previous research suggests that mis-specifying the distribution of covariates as multivariate normal in missing-data problems has negligible impact on regression parameter estimates [17]. Estimation of β proceeds by re-expressing it as $\beta=\sum^{-1}_{(xx)}\sum_{(xs)}$, where $\Sigma_{(xx)}$ is the $q\times q$ variance–covariance matrix of X and $\Sigma_{(xs)}$ is the $q\times 1$ covariance matrix of X with $Y_{(s)}$ [18]. The observed-data likelihood and estimator, $\hat{\beta}$, are provided in Appendix A. Appendix A also shows that $\rho^{(r)}_{(sp)}$ is only explicitly used in ML estimation of Model 3.

Estimates of β based on Model l are conditioned on λ_l, l=1,2, 3. Presenting multiple estimates as part of a sensitivity analysis treats all values of λ_l as exchangeable, although often some values are considered more plausible than others. Also, multiple estimates may not satisfy subject-matter scientists. One solution to both problems is to specify a distribution for λ_l, $f_{\lambda l}(\cdot)$, defined over a range of plausible values of λ_l, such that $f_{\lambda l}(\cdot|X)=f_{\lambda l}(\cdot)$. Integrating $\Sigma_{(xs)}$ over $f_{\lambda l}(\cdot)$ produces a single β that is a weighted average of λ_l-specific β's. Let $\mu_{\lambda l}$ denote the expected value of λ_l from $f_{\lambda l}(\cdot)$. Integrating over λ_l results in replacing λ_l with $\mu_{\lambda l}$ when specifying $\beta^{(0)}$.

2.3. Estimation: Multiple Imputation

Multiply imputing missing $Y_{(s)}$ involves a two-step procedure for creating each completed data set. Step 1 is to estimate the parameters ($\beta^{(1)}, \alpha^{(0)}, \sigma^{(1)}_{(ss)}, \sigma^{(0)}_{(pp)}$), and step 2 is to impute the missing data, conditional on parameter estimates [4]. We propose a normal imputation method that leads to an approximate Bayesian analysis (see Appendix A). Unlike maximum likelihood, Appendix A shows that $\rho^{(r)}_{(sp)}$ is used in estimation for Models 1–3 to impute missing $Y_{(s)}$ given observed $Y_{(p)}$. The MI algorithm in Appendix A is conditional on λ_l for Model l, l=1,2, 3. Adding a step where λ_l is simulated from $f_{\lambda_l}(\cdot)$ produces estimates that average over $f_{\lambda_l}(\cdot)$. The simulated λ_l is used to specify $\beta^{(0)}$ according to pattern-mixture restrictions. The data provide no information about λ_l, thus $f_{\lambda_l}(\cdot|Y_{(\text{obs})},X)=f_{\lambda_l}(\cdot)$.

3. Proxy–subject validation data

Until now, we have considered the study design in which only one of the subjects or proxies provides a response. In this section, we accommodate studies that include a random validation subsample from subjects with $R_{(s)}$=1, where both the subject and proxy provide responses. Several gerontologic studies have used this design to quantify proxy bias [5,6,19], however none have used validation data in analyses to correct for proxy bias.

We introduce new notation for validation data. Let $R_{(p)}$ indicate whether a proxy provides a response, where $R_{(p)}$=1 denotes observed $Y_{(p)}$, and $R_{(p)}$=0 denotes unobserved $Y_{(p)}$. Without validation data, $R_{(s)}=1-R_{(p)}$. However, with validation data, $R_{(s)}R_{(p)}=1$ indicates inclusion in the validation sample, and $R_{(s)}(1-R_{(p)})=1$ indicates exclusion from the validation sample. Validation proxies are only randomly selected among subjects with $R_{(s)}$=1. Let $\pi_{p|s}=\text{Prob}(R_{(p)}=1|R_{(s)}=1)$ be the investigator-defined probability of selection into the validation sample. Now, $Y_{(\text{obs})i}$ equals either ($Y_{(s)i}$, $Y_{(p)i}$), $Y_{(s)i}$, or $Y_{(p)i}$ depending on whether $R_{(s)i}R_{(p)i}=1$, $R_{(s)i}(1-R_{(p)i})=1$, or $(1-R_{(s)i})=1$, respectively.

An implication of using validation data is that previously unidentifiable parameters are now estimable, and weaker assumptions can be posited for parameters that remain unidentifiable. Given that selection into the validation sample is random and Y is assumed to be bivariate normal, $f(Y|X,R_{(s)}=1,R_{(p)}=r)=f(Y|X,R_{(s)}=1)$ for $r\in\{0,1\}$. Therefore, $\sigma^{(1)}_{(pp)}$ and $\sigma^{(1)}_{(sp)}$ can be estimated by observations in the validation sample, and equation (4) can be relaxed. In this case, equations (2) and (3) imply $\rho^{(0)}_{(sp)}=\rho^{(1)}_{(sp)}\sqrt{\sigma^{(1)}_{(pp)}/\sigma^{(0)}_{(pp)}}$. Let Model V($l$), l=2, 3, denote Model l with equation (4) relaxed. Validation data can help to inform a sensitivity analysis. When Model V(2) is posited, it is natural to treat $\lambda_2=0_q$ as an ‘anchor’ and to perform a sensitivity analysis around departures from the assumption $\beta^{(0)}=\alpha^{(0)}$. With validation data, one can estimate $\beta^{(1)}$ and $\alpha^{(1)}$. Thus, one can treat $\lambda_2=\beta^{(1)}-\alpha^{(1)}$ as the anchor and perform a sensitivity analysis around the assumption $\beta^{(0)}-\alpha^{(0)}=\beta^{(1)}-\alpha^{(1)}$. When Model V(3) is posited, $\lambda_3=0_q$ is also the natural anchor for sensitivity analysis, which implies that $Y_{(p)}$ is conditionally independent of X given $Y_{(s)}$ and $R_{(s)}$ (i.e. measurement error model). With validation data, one can estimate $\lambda_3=\alpha^{(1)}-\gamma\beta^{(1)}$. In this case, departures from equation (7) can be more easily considered, where $\lambda^{(0)}$ is an unidentifiable q-vector that can be anchored at $\lambda^{(1)}$. Estimation of these parameters using ML and MI via Gibbs sampling [20] is in Appendix B.

Another implication of validation data is that associations of proxy characteristics (e.g. age, relationship, and living arrangement with subject) with $Y_{(s)}$ can be identified when $R_{(s)}$=1. Thus, proxy characteristics can easily be included as auxiliary data. When Z is a categorical proxy characteristic, the analyst can estimate $E[Y_{(s)}-Y_{(p)}|X,Z=z,R_{(s)}=1]$ and

$E[Y_{(p)}|X,Y_{(s)},Z=z,R_{(s)}=1]$ to find Z-specific λ_2 and λ_3, respectively. When Z is continuous, a two-stage linear model can estimate Z-adjusted λ_2 or λ_3. In this case, $\beta^{(0)}$ is determined by relating $E[Y_{(d)}|Z, X,R_{(s)}R_{(p)}=1]$, $d\in\{p, s\}$ to $E[Y_{(s)}|Z, X,R_{(s)}=0]$.

4. Simulation studies

We performed two sets of simulation studies, one with and one without validation data, to compare the proposed ML and MI methods with two common alternatives: linear regression with only subject data (subjects requiring proxies excluded) and linear regression with proxy data substituted for missing subject data (i.e. single imputation). For each set of simulations, $N_{\text{sim}}=1000$ 'studies' were simulated consisting of either $n=100$ or 250 subjects. For each subject, $R_{(s)}$ was simulated from a Bernoulli distribution with $\pi_s=0.65$. Conditional on $R_{(s)}$, a covariate X_2 was simulated from a Bernoulli distribution with probability $0.4+0.2R_{(s)}$. Conditional on $R_{(s)}$ and X_2, a covariate X_1 was simulated from a normal distribution with mean $2.5+0.5X_2-R_{(s)}-0.25X_2R_{(s)}$ and variance $1.5-0.5R_{(s)}$. When $R_{(s)}=1$, $Y_{(s)}$ was simulated from a normal distribution with mean X_1+X_2, $\beta^{(1)}=(1, 1)$ and $\sigma^{(1)}_{(ss)}=1$. When $R_{(s)}=0$, $Y_{(p)}$ was simulated from a normal distribution with mean $0.5X_2+0.5X_1$, i.e. $\alpha^{(0)}=(0.5,0.5)$, and $\sigma^{(0)}_{(pp)}=1.5$. We specified equations (2)-(4) to be true, and considered two values of $\rho^{(r)}_{(sp)}$, 0.5 and 0.8. We estimated β for three true values of $\beta^{(0)}$: $\beta^{(1)}$, $\alpha^{(0)}$, and $\alpha^{(0)}+0.75$. When $\beta^{(0)}=\beta^{(1)}$, $\lambda_1=(0, 0)$, and $\beta=1.0$. When $\beta^{(0)}=\alpha^{(0)}$, $\lambda_2=(0, 0)$ and, when $\rho^{(r)}_{(sp)}=0.5$ or 0.8, $\lambda_3=(0.19,0.19)$ or $(0.01,0.01)$, respectively, resulting in $\beta=(0.61,1.19)$. Lastly, when $\beta^{(0)}=\alpha^{(0)}+0.75$, $\lambda_2=(0.75,0.75)$, and, when $\rho^{(r)}_{(sp)}=0.5$ or 0.8, $\lambda_3=(-0.26, -0.26)$ or $(-0.72, -0.72)$, respectively, resulting in $\beta=(1.34, 0.64)$. Assumed $\rho^{(r)}_{(sp)}$ is not used in subject only, subject+proxy, or ML estimation of Models 1 and 2. Observed $Y_{(p)}$ was used to simulate missing $Y_{(s)}$ in MI estimation of Models 1-3, which required specification of $\rho^{(r)}_{(sp)}$.

When validation data were included, the simulation procedure was modified. In particular, when $R_{(s)}=1$, $R_{(p)}$ was simulated from a Bernoulli distribution with $\pi_{p|s}=0.6$, and when $R_{(s)}R_{(p)}=1$, $(Y_{(s)},Y_{(p)})$ was simulated from a bivariate normal distribution with mean $(X\beta^{(1)},X\alpha^{(1)})$, and variance–covariance matrix $\Sigma^{(1)}$. When $\beta^{(0)}=\beta^{(1)}$ was specified, we set $\alpha^{(1)}=\beta^{(1)}$. Otherwise, when $\beta^{(0)}=\alpha^{(0)}$ or $\beta^{(0)}=\alpha^{(0)}+0.75$, $\alpha^{(1)}$ was specified so that $\beta^{(1)}-\alpha^{(1)}=\beta^{(0)}-\alpha^{(0)}$. For both simulation studies (with and without validation data), MI was performed by imputing 20 sets of missing $Y_{(s)}$. Standard errors for ML were estimated using 150 bootstrap samples.

4.1. Simulation results without validation data

When validation data were not simulated, λ_1, λ_2, and λ_3 were treated as fixed quantities. Table I shows that the proposed ML and MI methods produced negligible bias and good coverage when $\beta^{(0)}$ was correctly specified. Linear regression using only subject data performed well only when $\beta^{(0)}=\beta^{(1)}$. Also, linear regression using single imputation performed well only when $\beta^{(0)}=\alpha^{(0)}$. Bias and coverage for all methods, however, were sensitive to misspecification about $\beta^{(0)}$. For Models 1 and 2, using $Y_{(p)}$ to impute missing $Y_{(s)}$ provided no efficiency benefits over ML estimation. Estimates produced using Model 3 were less efficient than those produced using Model 2, because, when Model 3 is posited, $\hat{\beta}$ depends on $\widehat{\sigma}^{(1)}_{(ss)}$ and $\widehat{\sigma}^{(0)}_{(pp)}$. However, standard errors from Model 3 assuming $\rho^{(r)}_{(sp)}=0.8$ were smaller than those assuming $\rho^{(r)}_{(sp)}=0.5$. This is not surprising, because $\beta^{(0)}$ in Model 3 has an inverse relationship with $\rho^{(r)}_{(sp)}$. When there are no validation data, $\rho^{(r)}_{(sp)}$ is treated as a constant. Therefore, higher $\rho_{(sp)}^{(r)}$ in absolute value leads to lower variability of estimated $\beta^{(0)}$ and therefore lower variability of estimated β, because β is a weighted average of $\beta^{(0)}$ and $\beta^{(1)}$. Additionally, results using Model 3 under correct assumptions were more accurate when n=250 than when n=100.

To further investigate model and method performance for smaller sample sizes, we repeated the simulation study with n=50. The largest magnitudes of bias observed using ML on Models 2 and 3 were 6 and 10 per cent, respectively; and the largest magnitudes of bias observed using MI on Models 2 and 3 were 2 and −14 per cent, respectively.

4.2. Simulation results with validation data

The parameters λ_2 and λ_3 were estimated using the validation data. Table II shows that in most cases when $\beta^{(0)}$ was correctly specified, the proposed methods produced results with small bias and good coverage. However, when $\rho_{(sp)}^{(r)}=0.5$, both ML and MI estimation of Model V(3) produced results with large bias. Also, when n=100, MI estimation of Model V(2) produced some results with non-negligible bias. Bias and coverage were sensitive to misspecification about $\beta^{(0)}$. Standard errors for Models V(2) and V(3) were larger than those for Models 2 and 3, respectively, due to extra variability from estimating λ_2, λ_3, and $\sigma_{(sp)}^{(1)}$ versus plugging in analyst-specified fixed values. Unlike estimation of Model 2, standard errors from both ML and MI estimation of Model V(2) were smaller when $\rho_{(sp)}^{(r)}=0.8$ than when $\rho_{(sp)}^{(r)}=0.5$.

We also repeated the simulation study with n=50. The largest magnitude of bias observed using ML on Model V(2) was 3 per cent. When $\rho_{sp}^{(r)}=0.5$, biases over 100 per cent were observed using ML on Model V(3); however, when $\rho_{sp}^{(r)}=0.8$, the largest bias was 13 per cent. The largest magnitude of bias observed using MI on Model V(2) was 4 per cent. When $\rho_{sp}^{(r)}=0.5$, the largest bias observed using MI on Model V(3) was 24 per cent; however, when $\rho_{sp}^{(r)}=0.8$, the largest bias was 6 per cent.

5. Data application: BHS-2

We applied the proposed methods to BHS-2, a study of older adults' physical recovery from hip fracture. The analysis goal was to determine the relationship between patient sex and age at the time of fracture with disability for 12 months post-fracture, where disability was measured as the number of IADLs that the patient requires human or equipment assistance to perform. The scale (range: 0–7) consisted of seven tasks: using the telephone, managing money, managing medications, traveling to places outside of walking distance, shopping, preparing meals, and doing housework (see [19]). Analyses included 248 hip-fracture patients (41 men, 207 women) aged ≥65 years. Among 248 patients, 169 patients provided responses about IADLs, and proxies provided IADL reports for the other 79 patients, $\widehat{\pi}_s=\frac{169}{248}=0.68$. Among the 169 patients who provided self-reports, proxies for 91 patients also provided IADL reports, $\widehat{\pi}_{p|s}=\frac{91}{169}=0.54$. We performed two sets of analyses. The first analysis ignored validation data, thus $\sigma_{(sp)}^{(r)}$ (or $\rho_{(sp)}^{(r)}$), λ_2, and λ_3 were analyst-specified. In a previously published validation study of a different cohort of hip-fracture patients, Magaziner *et al.* [5] found a correlation of 0.70 between subject and proxy IADL reports.

Sex- and age-specific proxy bias has not been reported among hip-fracture patients. However, proxy bias has been reported for subgroups defined by other characteristics [5]. Thus, we specified assumptions about proxy bias for characteristics that are associated with age and sex. One characteristic associated with patient sex is living arrangement of the proxy with the patient. It has been shown that men tend to be younger than women at the time of fracture [21], and women have longer life expectancy than men. We presumed that proxies living with patients were often spouses, whereas proxies not living with patients were often offspring or unrelated care givers. Therefore, we expected that proxies living with patients were most often wives of male patients, whereas proxies not living with patients were most often offspring or unrelated care givers of female patients. Thus, bias from proxies who lived with the patient was thought to approximate proxy bias among male patients. Analogously, bias from proxies who did not live with the patient was thought to approximate proxy bias for female patients. Magaziner *et al.* [5] found that among patients living with proxies, patients reported an average of 0.49 fewer IADL dependencies than proxies, and among patients not living with proxies, patients reported an average of 0.23 fewer IADL dependencies than proxies. Thus, the magnitude of proxy bias was −0.49−(−0.23)=−0.26. Also, we presumed that proxies overreport patient IADLs compared with patients themselves at all ages, but that the magnitude of overreporting diminishes with older patient ages. The maximum value for IADL dependency was 7, thus the ceiling effect may limit the level of bias for the oldest patients. The ages of patients spanned 30 years (from 66 to 96 years) in this study. As an approximation, we specified that the degree of overreporting (bias) decreases by 0.01 IADL dependencies per year of age. We performed a sensitivity analysis where we assumed that λ_2=(0, 0), (−0.26,0.01), or (−0.52,0.02); where the second set were historical values derived from Magaziner *et al.* [5], and the third set is twice the historical values. We also assumed and $\rho^{(r)}_{(sp)}=0.70$ or 0.35, the historical value and half the historical value, respectively. MI for all three models depended on $\rho^{(r)}_{(sp)}$ to impute missing $Y_{(s)}$ using observed $Y_{(p)}$. In contrast, ML estimation depended on $\rho^{(r)}_{(sp)}$ only for Model 3. We estimated λ_3 by $\widehat{\alpha}^{(0)} - \rho^{(r)}_{(sp)}\sqrt{\widehat{\alpha}^{(0)}_{(pp)}(\widehat{\alpha}^{(1)}_{(ss)})^{-1}}(\widehat{\alpha}^{(0)}+\lambda_2)$. That is, the same sets of values for $\beta^{(0)}$ were assumed using both Models 2 and 3. We performed two other analyses, one assuming that $\beta^{(0)}=\beta^{(1)}$, and another assuming that λ_3=(0, 0) (i.e. the outcome measurement error model [16]). The second set of analyses included validation data by estimating λ_2 as $\hat{\beta}^{(1)}-\hat{\alpha}^{(1)}$ and estimating λ_3 as $\widehat{\alpha}^{(1)} - \widehat{\sigma}^{(1)}_{(sp)}/\widehat{\sigma}^{(1)}_{(ss)}\widehat{\beta}^{(1)}$.

Table III shows that, when validation data were excluded, the estimated coefficient for sex ranged from 0.76 to 3.34. The minimum was derived when MI was used with Model 2 assuming that λ_2=(−0.52,0.02) and $\rho^{(0)}$=0.35. The maximum was calculated with ML assuming Model 3 with λ_3=0 and $\rho^{(0)}_{(sp)}=0.35$. The estimated coefficient for age ranged from 0.10 using all methods assuming $\beta^{(0)}=\beta^{(1)}$ to 0.33 with ML assuming Model 3 and with λ_3=0 and $\rho^{(0)}_{(sp)}=0.35$. In absolute terms, the coefficient for age was more homogeneous than that for sex over the range of assumptions. This result is not surprising, because values of λ_2 for age were close to 0. In relative terms, however, both coefficients varied over the assumptions. For MI estimation of Model 3, assuming $\rho^{(0)}_{(sp)}=0.70$ produced smaller standard errors than assuming $\rho^{(0)}_{(sp)}=0.35$. Assumptions about $\rho^{(0)}_{(sp)}$ had a small effect on parameter estimates and standard errors for Models 1 and 2. When Model 2 was assumed, ML produced smaller standard errors than MI; however, the opposite was true when Model 3 was assumed. Analysis with validation data resulted in MLEs $\widehat{\rho}^{(1)}_{(sp)}=0.87$,

$\hat{\beta}^{(1)} - \hat{\alpha}^{(1)} = (0.52, 0.04)$, and $\hat{\alpha}^{(1)} - (\hat{\alpha}^{(1)}_{(sp)} / \hat{\alpha}^{(1)}_{(ss)}) \hat{\beta}^{(1)} = (-0.36, -0.03)$. Thus, assumptions using validation data differed from those derived from Magaziner *et al.* [5]. Table III shows that Models V(2) and V(3) resulted in estimated coefficients for sex of approximately 1.14, and estimated coefficients for age of 0.14.

When validation data were excluded, Models 2 and 3 were preferred because they were more flexible than other options considered. Model 2 using ML was advantageous because, unlike Model 3, it only depended on one sensitivity analysis parameter (λ_2), and produced smaller standard errors than the analogous model estimated using MI. When validation data were included, Model V(2) was preferred. Differences in estimates between Models V(2) and V(3) and between ML and MI were negligible, but Model V(2) estimated with ML produced the smallest standard errors. Qualitative conclusions from BHS-2 were robust to the assumptions considered: male sex and older age were associated with more IADL dependencies. However, the magnitude of association was sensitive to the assumptions examined here.

6. Discussion

This paper proposed methods based on pattern-mixture models to analyze normal data with proxy respondents. The methods were developed to handle studies both with and without validation subsamples of incompletely observed proxy respondents. The models proposed here are distinct from a recently published proxy pattern-mixture model where, unlike this paper, the authors used the term ‘proxy’ to refer to the function of completely observed covariates most predictive of the incompletely observed outcome [22].

Previous empirical studies of older adults have found evidence that proxy responses are systematically biased compared with subject responses [5-8]. Despite these findings, proxy data are often substituted for missing subject data without considering the implied assumptions. In contrast, our proposed methods involve explicating assumptions about missing subject data. The analyst can relate the distribution of missing subject data to identifiable distributions for observed proxy or subject data.

Although proxy data can also be treated as a source of outcome measurement error [16], conceptualizing the problem instead as one of missing data is beneficial. In the measurement error framework, it is common to assume that proxy data are surrogates for patient data (i.e. $\lambda_3 = 0_q$ in Model 3), in other words that measurement error is nondifferential. A benefit of our models is that they can easily handle differential measurement error with respect to covariates and auxiliary variables in the analysis model. Also, standard methods for measurement error focus on the scenario where $Y_{(p)}$ is observed for all subjects, and $Y_{(s)}$ may be observed for a random validation subsample; i.e. no selection bias. Our proposed models handle selection bias in the proxy-data problem, because the distribution of $Y_{(s)}$ may differ between those with $Y_{(s)}$ observed and those for whom only $Y_{(p)}$ is observed.

Simulation studies showed that, in general, our proposed methods produce results with low bias and good coverage when proxy bias is correctly specified. However, some caveats should be kept in mind. Estimation with Model V(3) using validation data is only advisable when the proxy and subject responses are highly correlated conditional on covariates; and MI estimation of Model V(2) is only advisable with large samples. These findings illustrate that while validation data can be a valuable resource, if it is of low quality and quantity, it can negatively impact model performance. Also, the simulations showed that when sample sizes are as small as 50, Models 2 and V(2) are generally more reliable than Models 3 and V(3), respectively, because $\beta^{(0)}$ in Models 3 and V(3) depend multiplicatively on inverse variances or covariances.

Despite these caveats, our approach provides more flexibility in performing sensitivity analyses about proxy bias than standard *ad hoc* methods such as analyzing only subject data or singly imputing proxy data for missing subject data. The BHS-2 data analysis highlights the benefit of validation proxies, because including validation proxy data allows weaker assumptions to be made about parameters. However, the proposed models only handle proxy bias for normal outcomes. Future research involves extending the methods for non-normal outcomes and addressing proxy bias in covariates. The missing-data framework will ease extensions for handling additional missing data (i.e. when responses from subjects and proxies are both not available). Wang *et al.* [23] developed a selection model to simultaneously handle missing data and measurement error; however, this model was based on the surrogacy assumption and does not consider the scenario where only one of the subject or proxy response is observed.

Lastly, although our approach facilitates a sensitivity analysis about the magnitude of proxy bias among those with missing $Y_{(s)}$, it may be difficult to determine a range of plausible values for unidentifiable parameters. Validation data are particularly beneficial here, because they can help provide plausible anchors for these parameters. When validation data are not available, historical validation studies, such as those used in the BHS-2 example, can provide initial assumptions. Otherwise, subject-matter experts are generally regarded as the best source of information for sensitivity analyses [24].

Appendix A: Estimation without validation data

A.1. Maximum likelihood

Let n_1 be the number with $R_{(s)}$=1, and denote $n_0=n-n_1$. Let Prob($R_{(s)}$=1)=π_s, and $\phi_1=(\beta^{(1)}, \alpha^{(0)}, \sigma^{(1)}_{(ss)}, \sigma^{(0)}_{(pp)}, \mu^{(0)}_{(x)}, \sum^{(0)}_{(xx)}, \mu^{(1)}_{(x)}, \sum^{(1)}_{(xx)}, \pi_s)$. The observed-data likelihood is

$$L(\phi_1;Y_{(\text{obs})},X,R_{(s)})\alpha\pi_s^{n_1}(1-\pi_s)^{n_0}\prod_i f(X_i|R_{(s)i})f(Y_{(s)i}|X_i,R_{(s)i})^{R_{(s)i}}f(Y_{(p)i}|X_i,R_{(s)i})^{(1-R_{(s)i})}, \tag{A1}$$

where $f\,(Y_{(d)i}|X_i, R_{(s)i})$ is normal, $d\in\{p, s\}$ and $f\,(X_i|R_{(s)i})$ is multivariate normal with mean $\mu^{(r)}_{(x)}$and variance–covariance $\sum^{(r)}_{(xx)}$. Plug MLEs from equation (A1) into $\Sigma_{(xs)}$:

$$\sum\nolimits_{(xs)}=\sum_{r\in\{0,1\}}\pi_s^r(1-\pi_s)^{1-r}[\sum\nolimits^{(r)}_{(xx)}\beta^{(r)}+(\mu^{(r)}_{(x)}-\mu_{(x)})^{'}(\mu^{(r)}_{(x)}\beta^{(r)}-\mu_{(y_{(s)})})], \tag{A2}$$

where $\mu_{(x)}=\pi_s\mu^{(1)}_{(x)}+(1-\pi_s)\mu^{(0)}_{(x)}$ and $\mu_{(Y_{(s)})}=\pi_s\mu^{(1)}_{(x)}\beta^{(1)}+(1-\pi_s)\mu^{(0)}_{(x)}\beta^{(0)}$.

A.2. Multiple imputation

First, draw $\sigma^{(1)}_{(ss)}$ from $S^{(1)}_{(ss)}(n_1-q)/\chi^2_{n_1-q}$, where $S^{(1)}_{(ss)}$ is the mean-squared error from regressing $Y_{(s)}$ on X among those with $R_{(s)}$=1, and χ^2_η is a chi-square random variable with η degrees of freedom. Second, draw $\beta^{(1)}$ from $\text{MVN}(\widehat{\beta}^{(1)}, \sigma^{(1)}_{(ss)}(n_1\widehat{\sum}^{(1)}_{(xx)})^{-1}|R_{(s)}=1)$. Next, draw $\sigma^{(0)}_{(pp)}$ from $S^{(0)}_{(pp)}(n_0-q)/\chi^2_{n_0-q}$, where $S^{(0)}_{(pp)}$ is the mean-squared error from regressing $Y_{(p)}$ on

X among those with $R_{(s)}=0$. Then, draw $\alpha^{(0)}$ from $\mathrm{MVN}(\widehat{\alpha}^{(0)}, \sigma^{(0)}_{(pp)}(n_0 \widehat{\sum}^{(0)}_{(xx)})^{-1} | R_{(s)}=0)$. Use simulated parameters and assumed model to find $\beta^{(0)}$. Simulate missing Y_s from a normal distribution with mean $X\beta^{(0)}+\rho^{(0)}_{(sp)}\sqrt{\sigma^{(0)}_{(ss)}/\sigma^{(0)}_{(pp)}}(Y_{(p)} - X\alpha^{(0)})$ and variance $\sigma^{(0)}_{(ss)}(1 - \rho^{(0)^2}_{(sp)})$. Repeat the steps M times to create M completed data sets, and calculate final parameter and standard error estimates using the usual combining rules [4].

Appendix B: Estimation with validation data

B.1. Maximum likelihood

Let n_{11} be the number in the validation sample, let $n_{10}=n_1-n_{11}$ be the number with $R_{(s)}=1$ not selected into the validation sample, and let $\phi_2=(\phi_1, \alpha^{(1)}, \sigma^{(1)}_{(pp)}, \sigma^{(1)}_{(sp)}, \pi_{p|s})$. The observed-data likelihood is

$$L(\phi_2;Y_{(\mathrm{obs})}, X, R_{(s)}, R_{(p)}) \propto (\pi_s\pi_{p|s})^{n_{11}}[\pi_s(1 - \pi_{p|s})]^{n_{10}}(1 - \pi_s)^{n_0}\prod_i f(X_i|R_{(s)i}) \times [f(Y_{(s)i}|X_i, R_{(s)i})^{(1-R_{(p)i})} f(Y_{(s)i}, Y_{(p)i}|X_i, R_{(s)i})^{R_{(p)i}}]^{R_{(s)i}} f(Y_{(p)i}|X_i, R_{(s)i})^{(1-R_{(s)i})}. \tag{B1}$$

Plug MLEs from equation (B1) into equation (A2) to estimate β.

B.2. Multiple imputation

Draw M-independent simulations of $(\alpha^{(0)}, \sigma^{(0)}_{(pp)})$ as described in Appendix A. Use Gibbs sampling to simulate $(\beta^{(1)}, \alpha^{(1)}, \Sigma^{(1)})$, because more observations are available to estimate $\beta^{(1)}$ and $\sigma^{(1)}_{(ss)}$ than that are available to estimate $\alpha^{(1)}$, $\sigma^{(1)}_{(sp)}$, and $\sigma^{(1)}_{(pp)}$, complicating the posterior distributions [20]. When $R_{(s)}=1$, treat $Y_{(p)}$ as data missing according to a known mechanism. Let $(\beta^{(1)}, \alpha^{(1)})$ have priors proportional to a constant that are independent of $\Sigma^{(1)^{-1}}$, which has a Wishart prior, $W(v,A)$, with v degrees of freedom, and 2×2 symmetric positive-definite prior precision matrix A. Y_i is a 2-vector, thus $v \geq 2$. First, specify initial values for the $Y_{(p)}$ where $R_{(s)}(1-R_{(p)})=1$. Let $Y^{comp(j)}_{(p)}$ denote the vector of completed proxy data (observed and imputed) at the jth iteration, where $Y^{\mathrm{comp}(j)}_{(p)i}=Y_{(p)i}$ if $R_{(s)i}R_{(p)i}=1$. Continue iteration j by simulating $(\beta^{(1j)},\alpha^{(1j)}, \Sigma^{(1j)})$ from

$$\sum^{(1j)^{-1}} | Y_{(s)}, Y^{comp(j)}_{(p)}, X, R_{(s)}=1 \sim W(v+n_1 - 1, (e^{(j)\prime} e^{(j)}+A^{-1})^{-1})$$

$$\beta^{(1j)}, \alpha^{(1j)} | \sum^{(1j)}, Y_{(s)}, Y^{\mathrm{comp}(j)}_{(p)}, X, R_{(s)}=1 \sim \mathrm{MVN}((\widehat{\alpha}^{(1j)},\widehat{\beta}^{(1)}), \sum^{(1j)} \otimes (n_1 \widehat{\sum}^{(1)}_{(xx)})^{-1}),$$

where $e^{(j)}$ is n_1×2 with ith row $(Y_{(s)i} - X_i\widehat{\beta}^{(1)}, Y^{\mathrm{comp}(j)}_{(p)i} - X_i\widehat{\alpha}^{(1j)})$, $\hat{\beta}^{(1)}$ is MLE of $\beta^{(1)}$, and $\hat{\alpha}^{(1j)}$ is MLE of $\alpha^{(1)}$ using $Y^{\mathrm{comp}(j)}_{(p)}$. Begin iteration j+1 by drawing $Y^{\mathrm{comp}(j+1)}_{(p)}$ from

$$Y_{(p)i}^{\mathrm{comp}(j+1)} | \beta^{(1j)}, \alpha^{(1j)}, \sum^{(1j)}, Y_{(s)i}, X_i, R_{(s)i}(1-R_{(p)i})=1=N\left(X_i\alpha^{(1j)}+\frac{\sigma_{(sp)}^{(1j)}}{\sigma_{(ss)}^{(1j)}}(Y_{(s)i}-X_i\beta^{(1j)}, \sigma_{(pp)}^{(1j)}-\frac{\sigma_{(sp)}^{(1j)^2}}{\underset{(ss)}{(1j)}}\right).$$

After completing the Gibbs sampler, select M draws of the simulated parameters spaced far enough apart between iterations to avoid autocorrelation. Set $\beta^{(0)}$ according to the assumed model, and draw M independent sets of $Y_{(s)}$ as in Appendix A

Acknowledgments

Contract/grant sponsor: National Institute of Health; contract/grant numbers: K12HD043489, K12HD055931, K23AG027746, R37AG009901, R01AG09902

References

1. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist 1969;9:179–186. [PubMed: 5349366]
2. Magaziner, J. The use of proxy respondents in health studies of the aged. In: Wallace, RB.; Woolson, RF., editors. The Epidemiologic Study of the Elderly. Oxford University Press; New York: 1992. p. 120-129.
3. Moinpour CM, Lyons B, Grevstad PK, Lovato LC, Crowley J, Czaplicki K, Buckner ZM, Ganz PA, Kelly K, Gandara DR. Quality of life in advanced non-small-cell lung cancer: results of a southwest oncology group randomized trial. Quality of Life Research 2002;11:115–126.10.1023/A: 1015048908822 [PubMed: 12018735]
4. Rubin, DB. Multiple Imputation for Non-response in Surveys. Wiley; New York: 1987.
5. Magaziner J, Simonsick EM, Kashner TM, Hebel JR. Patient-proxy response comparability on measures of patient health and functional status. Journal of Clinical Epidemiology 1988;41:1065–1074. [PubMed: 3204417]
6. Magaziner J, Bassett SS, Hebel JR, Gruber-Baldini A. Use of proxies to measure health and functional status in epidemiologic studies of community-dwelling women aged 65 years and older. American Journal of Epidemiology 1996;143:283–292. [PubMed: 8561163]
7. Neumann PJ, Araki SS, Gutterman EM. The use of proxy respondents in studies of older adults: lessons, challenges, and opportunities. Journal of the American Geriatrics Society 2000;48:1646–1654. [PubMed: 11129756]
8. Snow AL, Cook KF, Lin PS, Morgan RO, Magaziner J. Proxies and other external raters: methodological considerations. Health Services Research 2005;40:1676–1693.10.1111/j.1475–6773.2005.00447.x [PubMed: 16179002]
9. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. Statistics in Medicine 1989;8:431–440. [PubMed: 2727467]
10. Huang R, Liang Y, Carriere KC. The role of proxy information in missing data analysis. Statistical Methods in Medical Research 2005;14:457–471.10.1191/0962280205sm411oa [PubMed: 16248348]
11. Rubin DB. Inference and missing data. Biometrika 1976;63:581–592.
12. Baker SG, Kramer BS. A perfect correlate does not a surrogate make. BMC Medical Research Methodology 2003;3:16.10.1186/1471-2288-3-16 [PubMed: 12962545]
13. Little RJA. Pattern-mixture models for multivariate incomplete data. Journal of the American Statistical Association 1993;88:125–134.
14. Little RJA. A class of pattern-mixture models for normal incomplete data. Biometrika 1994;81:471–483.
15. Little RJA, Wang Y. Pattern-mixture models for multivariate incomplete data with covariates. Biometrics 1996;52:98–111. [PubMed: 8934587]

NIH-PA Author Manuscript
NIH-PA Author Manuscript

16. Buonaccorsi JP. Measurement error in the response in the general linear model. Journal of the American Statistical Association 1996;91:633–642.

17. Parzen M, Lipsitz SR, Ibrahim JG, Lipshultz S. A weighted estimating equation for linear regression with missing covariate data. Statistics in Medicine 2002;21:2421–2436.10.1002/sim. 1195 [PubMed: 12210626]

18. Glynn RJ, Laird NM, Rubin DB. Multiple imputation in mixture models for nonignorable nonresponse with follow-ups. Journal of the American Statistical Association 1993;88:984–993.

19. Magaziner J, Zimmerman SI, Gruber-Baldini AL, Hebel JR, Fox KM. Proxy reporting in five areas of functional status. American Journal of Epidemiology 1997;146:418–428. [PubMed: 9290502]

20. Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the bayesian restoration of images. IEEE Transactions in Pattern Analysis and Machine Intelligence 1984;6:721–741.

21. Hawkes WG, Wehren L, Orwig D, Hebel JR, Magaziner J. Gender differences in functioning after hip fracture. Journals of Gerontology Series A—Biological Sciences and Medical Sciences 2006;61:495–499.

22. Andridge, RR.; Little, RJA. American Statistical Association Proceedings of the Survey Research Methods Section. Denver, CO: 2008. Proxy pattern-mixture analysis for survey nonresponse; p. 3261-3268.

23. Wang CY, Huang Y, Chao EC, Jeffcoat MK. Expected estimating equations for missing data, measurement error, and misclassification, with application to longitudinal nonignorable missing data. Biometrics 2008;64:85–95.10.1111/j.1541–0420.2007.00839.x [PubMed: 17608787]

24. White IR, Carpenter J, Evans S, Schroter S. Eliciting and using expert opinions about dropout bias in randomized controlled trials. Clinical Trials 2007;4:125–139.10.1177/1740774507077849 [PubMed: 17456512]

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table I

Simulation results without validation data (N_{sim} =1000,N_{bs} = 150); per cent bias of $\hat{\beta}$, $100(\hat{\beta}-\beta)/|\beta|$ (Bias), standard error for $\hat{\beta}$ (SE), empirical standard error for $\hat{\beta}$ (ESE), and 95 per cent confidence interval coverage per cent (95 per cent CI Cov.)

					β_1 (continuous)				β_2(binary)			
n	**True $\beta^{(0)}$**	**Assumed $\beta^{(0)}$**	**Method**	**Assumed $\rho^{(r)}_{(sp)}$**	**Bias**	**SE**	**ESE**	**95 per cent CI Cov.**	**Bias**	**SE**	**ESE**	**95 per cent CI Cov.**
100	$\beta^{(1)}$	$\beta^{(1)}$	Subject Only	—	<1	0.127	0.127	95.2	<1	0.274	0.275	94.6
		$\alpha^{(0)}$	Subject+Proxy	—	−39	0.109	0.117	7.6	19	0.259	0.255	89.2
		$\beta^{(1)}$	ML-Model 1	—	<1	0.127	0.128	94.7	<1	0.270	0.273	94.6
		$\beta^{(1)}$	MI-Model 1	0.5	<1	0.130	0.137	94.0	<1	0.279	0.279	94.5
		$\beta^{(1)}$	MI-Model 1	0.8	<1	0.130	0.131	94.0	<1	0.279	0.274	95.3
		$\alpha^{(0)}$	ML-Model 2	—	−39	0.115	0.117	9.6	19	0.251	0.259	88.3
		$\alpha^{(0)}$	MI-Model 2	0.5	−38	0.126	0.121	13.3	19	0.289	0.263	91.5
		$\alpha^{(0)}$	MI-Model 2	0.8	−38	0.114	0.120	9.5	19	0.267	0.256	89.7
		$\alpha^{(0)}$+0.75	ML-Model 2	—	33	0.118	0.119	22.7	−39	0.271	0.274	69.4
		$\alpha^{(0)}$+0.75	MI-Model 2	0.5	34	0.130	0.123	25.3	−35	0.300	0.278	79.2
		$\alpha^{(0)}$+0.75	MI-Model 2	0.8	35	0.120	0.116	17.8	−36	0.279	0.264	75.8
		$\alpha^{(0)}$	ML-Model 3	0.5	−37	0.173	0.172	42.8	17	0.341	0.349	91.1
		$\alpha^{(0)}$	ML-Model 3	0.8	−37	0.144	0.148	27.8	16	0.269	0.278	90.5
		$\alpha^{(0)}$	MI-Model 3	0.5	−36	0.171	0.179	44.9	17	0.349	0.351	91.6
		$\alpha^{(0)}$	MI-Model 3	0.8	−36	0.139	0.147	29.7	18	0.274	0.270	90.5
		$\alpha^{(0)}$+0.75	ML-Model 3	0.5	37	0.263	0.258	80.1	−44	0.414	0.402	87.3
		$\alpha^{(0)}$+0.75	ML-Model 3	0.8	40	0.251	0.243	73.9	−43	0.360	0.336	85.0
		$\alpha^{(0)}$+0.75	MI-Model 3	0.5	39	0.257	0.266	79.9	−39	0.406	0.396	89.7
		$\alpha^{(0)}$+0.75	MI-Model 3	0.8	41	0.241	0.245	73.7	−40	0.347	0.322	87.0
100	$\alpha^{(0)}$	$\beta^{(1)}$	Subject Only	—	63	0.127	0.128	14.4	−16	0.274	0.272	89.9
		$\alpha^{(0)}$	Subject+Proxy	—	<1	0.109	0.119	92.3	<1	0.258	0.256	95.0
		$\beta^{(1)}$	ML-Model 1	—	62	0.127	0.129	16.9	−17	0.271	0.274	87.7
		$\beta^{(1)}$	MI-Model 1	0.5	63	0.127	0.133	18.3	−17	0.278	0.285	88.1
		$\beta^{(1)}$	MI-Model 1	0.8	63	0.129	0.139	20.4	−17	0.277	0.280	89.0
		$\alpha^{(0)}$	ML-Model 2	—	<1	0.116	0.119	93.5	<1	0.250	0.256	94.0

n	True $\beta^{(0)}$	Assumed $\beta^{(0)}$	Method	Assumed $\rho^{(r)}_{(sp)}$	β_1 (continuous) Bias	SE	ESE	95 per cent CI Cov.	β_2(binary) Bias	SE	ESE	95 per cent CI Cov.
		$\alpha^{(0)}$	MI-Model 2	0.5	<1	0.126	0.119	96.3	<1	0.287	0.259	96.8
		$\alpha^{(0)}$	MI-Model 2	0.8	<1	0.114	0.120	93.6	<1	0.267	0.271	94.3
		$\alpha^{(0)}+0.75$	ML-Model 2	—	118	0.118	0.117	0.0	−47	0.270	0.275	45.6
		$\alpha^{(0)}+0.75$	MI-Model 2	0.5	120	0.130	0.124	0.0	−47	0.300	0.272	54.1
		$\alpha^{(0)}+0.75$	MI-Model 2	0.8	120	0.121	0.124	0.0	−45	0.280	0.271	49.7
		$\alpha^{(0)}$	ML-Model 3	0.5	3	0.175	0.176	94.6	−1	0.342	0.334	94.9
		$\alpha^{(0)}$	ML-Model 3	0.8	2	0.147	0.147	95.0	−1	0.267	0.270	94.5
		$\alpha^{(0)}$	MI-Model 3	0.5	4	0.170	0.177	95.5	−1	0.346	0.339	95.8
		$\alpha^{(0)}$	MI-Model 3	0.8	4	0.141	0.156	92.1	−2	0.274	0.285	93.2
		$\alpha^{(0)}+0.75$	ML-Model 3	0.5	128	0.268	0.259	7.4	−51	0.417	0.401	74.8
		$\alpha^{(0)}+0.75$	ML-Model 3	0.8	127	0.248	0.229	2.5	−51	0.358	0.333	64.1
		$\alpha^{(0)}+0.75$	MI-Model 3	0.5	129	0.257	0.272	11.0	−51	0.407	0.408	73.6
		$\alpha^{(0)}+0.75$	MI-Model 3	0.8	129	0.242	0.244	4.8	−49	0.346	0.326	64.2
100	$\alpha^{(0)}+0.75$	$\beta^{(1)}$	Subject Only	—	−26	0.127	0.129	22.8	57	0.275	0.276	73.4
		$\alpha^{(0)}$	Subject+Proxy	—	−54	0.109	0.117	0.0	87	0.258	0.252	42.3
		$\beta^{(1)}$	ML-Model 1	—	−26	0.127	0.128	20.7	54	0.272	0.278	74.0
		$\beta^{(1)}$	MI-Model 1	0.5	−25	0.128	0.135	30.0	55	0.276	0.291	75.4
		$\beta^{(1)}$	MI-Model 1	0.8	−26	0.130	0.134	28.7	57	0.277	0.286	73.6
		$\alpha^{(0)}$	ML-Model 2	—	−55	0.116	0.119	0.0	84	0.250	0.252	43.2
		$\alpha^{(0)}$	MI-Model 2	0.5	−54	0.125	0.118	0.1	87	0.290	0.262	52.0
		$\alpha^{(0)}$	MI-Model 2	0.8	−54	0.114	0.123	0.0	86	0.267	0.261	46.7
		$\alpha^{(0)}+0.75$	ML-Model 2	—	<1	0.118	0.115	94.7	<1	0.271	0.266	95.2
		$\alpha^{(0)}+0.75$	MI-Model 2	0.5	<1	0.130	0.121	97.2	1	0.303	0.271	97.3
		$\alpha^{(0)}+0.75$	MI-Model 2	0.8	<1	0.120	0.116	95.7	2	0.279	0.281	94.4
		$\alpha^{(0)}$	ML-Model 3	0.5	−53	0.176	0.177	3.2	83	0.342	0.347	66.1
		$\alpha^{(0)}$	ML-Model 3	0.8	−53	0.144	0.146	0.1	80	0.268	0.261	50.5
		$\alpha^{(0)}$	MI-Model 3	0.5	−53	0.170	0.176	5.0	84	0.347	0.353	65.6
		$\alpha^{(0)}$	MI-Model 3	0.8	−52	0.139	0.152	0.9	84	0.273	0.273	51.4

					β_1 (continuous)				β_2(binary)			
n	**True $\beta^{(0)}$**	**Assumed $\beta^{(0)}$**	**Method**	**Assumed $\rho^{(r)}_{(sp)}$**	**Bias**	**SE**	**ESE**	**95 per cent CI Cov.**	**Bias**	**SE**	**ESE**	**95 per cent CI Cov.**
		$\alpha^{(0)}$+0.75	ML-Model 3	0.5	3	0.268	0.258	95.5	−6	0.417	0.387	97.2
		$\alpha^{(0)}$+0.75	ML-Model 3	0.8	2	0.244	0.216	96.3	−5	0.352	0.320	97.1
		$\alpha^{(0)}$+0.75	MI-Model 3	0.5	5	0.261	0.271	95.4	−6	0.412	0.410	97.3
		$\alpha^{(0)}$+0.75	MI-Model 3	0.8	3	0.237	0.241	95.7	−4	0.341	0.347	95.5
250	$\beta^{(1)}$	$\beta^{(1)}$	Subject Only	—	<1	0.079	0.082	94.1	<1	0.172	0.172	94.5
		$\alpha^{(0)}$	Subject+Proxy	—	−39	0.069	0.073	0.1	19	0.163	0.160	79.5
		$\beta^{(1)}$	ML-Model 1	—	<1	0.080	0.080	94.6	<1	0.171	0.175	94.3
		$\beta^{(1)}$	MI-Model 1	0.5	<1	0.081	0.081	95.5	<1	0.176	0.173	96.1
		$\beta^{(1)}$	MI-Model 1	0.8	<1	0.081	0.080	95.8	1	0.175	0.182	94.2
		$\alpha^{(0)}$	ML-Model 2	—	−39	0.074	0.076	0.0	17	0.159	0.159	80.2
		$\alpha^{(0)}$	MI-Model 2	0.5	−39	0.079	0.078	0.1	20	0.180	0.167	81.1
		$\alpha^{(0)}$	MI-Model 2	0.8	−39	0.071	0.076	0.0	20	0.166	0.163	79.5
		$\alpha^{(0)}$+0.75	ML-Model 2	—	34	0.073	0.072	0.4	−36	0.170	0.170	41.5
		$\alpha^{(0)}$+0.75	MI-Model 2	0.5	34	0.081	0.076	1.2	−36	0.188	0.172	53.4
		$\alpha^{(0)}$+0.75	MI-Model 2	0.8	34	0.074	0.076	0.5	−35	0.174	0.170	47.1
		$\alpha^{(0)}$	ML-Model 3	0.5	−37	0.107	0.111	7.8	15	0.209	0.207	87.7
		$\alpha^{(0)}$	ML-Model 3	0.8	−38	0.088	0.086	0.6	17	0.165	0.167	81.0
		$\alpha^{(0)}$	MI-Model 3	0.5	−39	0.104	0.109	9.1	19	0.216	0.216	84.1
		$\alpha^{(0)}$	MI-Model 3	0.8	−38	0.085	0.090	2.1	19	0.170	0.168	79.6
		$\alpha^{(0)}$+0.75	ML-Model 3	0.5	37	0.154	0.159	31.9	−40	0.241	0.236	63.3
		$\alpha^{(0)}$+0.75	ML-Model 3	0.8	36	0.140	0.141	21.9	−38	0.202	0.219	53.9
		$\alpha^{(0)}$+0.75	MI-Model 3	0.5	37	0.156	0.161	42.0	−38	0.242	0.244	70.6
		$\alpha^{(0)}$+0.75	MI-Model 3	0.8	36	0.141	0.141	29.9	−36	0.200	0.196	58.2
250	$\alpha^{(0)}$	$\beta^{(1)}$	Subject Only	—	63	0.079	0.081	0.4	−16	0.172	0.172	80.7
		$\alpha^{(0)}$	Subject+Proxy	—	<1	0.068	0.073	93.8	<1	0.162	0.160	95.7
		$\beta^{(1)}$	ML-Model 1	—	63	0.079	0.079	0.4	−17	0.171	0.170	78.3
		$\beta^{(1)}$	MI-Model 1	0.5	64	0.081	0.083	0.5	−16	0.174	0.175	81.8
		$\beta^{(1)}$	MI-Model 1	0.8	64	0.081	0.082	0.7	−16	0.175	0.174	81.1

NIH-PA Author Manuscript
NIH-PA Author Manuscript

n	True $\beta^{(0)}$	Assumed $\beta^{(0)}$	Method	Assumed $\rho_{(sp)}^{(r)}$	β_1 (continuous) Bias	SE	ESE	95 per cent CI Cov.	β_2(binary) Bias	SE	ESE	95 per cent CI Cov.
		$\alpha^{(0)}$	ML-Model 2	—	<1	0.074	0.075	93.8	<1	0.159	0.160	94.1
		$\alpha^{(0)}$	MI-Model 2	0.5	<1	0.078	0.073	96.9	<1	0.180	0.170	96.2
		$\alpha^{(0)}$	MI-Model 2	0.8	<1	0.071	0.075	94.0	<1	0.166	0.161	96.2
		$\alpha^{(0)}+0.75$	ML-Model 2	—	119	0.073	0.072	0.0	−47	0.170	0.169	9.5
		$\alpha^{(0)}+0.75$	MI-Model 2	0.5	120	0.081	0.077	0.0	−46	0.188	0.170	16.2
		$\alpha^{(0)}+0.75$	MI-Model 2	0.8	119	0.074	0.073	0.0	−46	0.174	0.166	11.5
		$\alpha^{(0)}$	ML-Model 3	0.5	2	0.106	0.108	93.9	<1	0.209	0.214	95.2
		$\alpha^{(0)}$	ML-Model 3	0.8	<1	0.087	0.087	94.5	<1	0.165	0.161	94.1
		$\alpha^{(0)}$	MI-Model 3	0.5	1	0.104	0.108	93.8	<1	0.215	0.225	94.0
		$\alpha^{(0)}$	MI-Model 3	0.8	<1	0.085	0.087	94.7	<1	0.170	0.166	95.3
		$\alpha^{(0)}+0.75$	ML-Model 3	0.5	122	0.152	0.148	0.0	−48	0.238	0.236	31.3
		$\alpha^{(0)}+0.75$	ML-Model 3	0.8	122	0.141	0.135	0.0	−48	0.201	0.186	14.5
		$\alpha^{(0)}+0.75$	MI-Model 3	0.5	122	0.153	0.159	0.0	−47	0.240	0.236	37.5
		$\alpha^{(0)}+0.75$	MI-Model 3	0.8	122	0.142	0.142	0.0	−47	0.202	0.190	17.3
250	$\alpha^{(0)}+0.75$	$\beta^{(1)}$	Subject Only	—	−26	0.079	0.080	1.2	57	0.172	0.170	43.4
		$\alpha^{(0)}$	Subject+Proxy	—	−54	0.069	0.074	0.0	86	0.163	0.160	7.5
		$\beta^{(1)}$	ML-Model 1	—	−26	0.079	0.079	0.9	54	0.171	0.171	47.0
		$\beta^{(1)}$	MI-Model 1	0.5	−26	0.081	0.081	3.1	57	0.175	0.174	48.2
		$\beta^{(1)}$	MI-Model 1	0.8	−26	0.081	0.083	1.6	58	0.175	0.173	45.2
		$\alpha^{(0)}$	ML-Model 2	—	−55	0.074	0.075	0.0	85	0.159	0.160	8.0
		$\alpha^{(0)}$	MI-Model 2	0.5	−55	0.079	0.076	0.0	86	0.181	0.162	12.6
		$\alpha^{(0)}$	MI-Model 2	0.8	−54	0.071	0.075	0.0	87	0.166	0.158	9.0
		$\alpha^{(0)}+0.75$	ML-Model 2	—	<1	0.073	0.073	94.7	<1	0.170	0.174	94.1
		$\alpha^{(0)}+0.75$	MI-Model 2	0.5	<1	0.082	0.076	96.1	1	0.189	0.177	96.3
		$\alpha^{(0)}+0.75$	MI-Model 2	0.8	<1	0.074	0.075	94.3	<1	0.174	0.174	94.6
		$\alpha^{(0)}$	ML-Model 3	0.5	−55	0.106	0.105	0.1	85	0.209	0.212	26.4
		$\alpha^{(0)}$	ML-Model 3	0.8	−54	0.088	0.087	0.0	83	0.165	0.163	11.4
		$\alpha^{(0)}$	MI-Model 3	0.5	−54	0.105	0.110	0.0	86	0.216	0.216	29.5

n	True $\beta^{(0)}$	Assumed $\beta^{(0)}$	Method	Assumed $\rho^{(r)}_{(sp)}$	β_1 (continuous)				β_2(binary)			
					Bias	SE	ESE	95 per cent CI Cov.	Bias	SE	ESE	95 per cent CI Cov.
		$\alpha^{(0)}$	MI-Model 3	0.8	−54	0.085	0.091	0.0	86	0.170	0.162	10.5
		$\alpha^{(0)}+0.75$	ML-Model 3	0.5	1	0.153	0.147	95.3	−1	0.241	0.246	94.2
		$\alpha^{(0)}+0.75$	ML-Model 3	0.8	1	0.140	0.143	94.2	−2	0.201	0.197	96.0
		$\alpha^{(0)}+0.75$	MI-Model 3	0.5	1	0.153	0.158	95.0	−1	0.241	0.241	95.2
		$\alpha^{(0)}+0.75$	MI-Model 3	0.8	2	0.143	0.142	95.4	−3	0.204	0.204	95.8

‘Subject Only’ refers to linear regression with only observed subject data, ‘Subject + Proxy’ refers to linear regression where proxy data substitutes for missing subject data, ML=maximum likelihood, MI=multiple imputation.

Table II

Simulation results with validation data (N_{sim}=1000,N_{bs}=150); per cent bias of $\hat{\beta}$, 100($\hat{\beta}$−β)/|β| (Bias), standard error for $\hat{\beta}$ (SE), empirical standard error for $\hat{\beta}$ (ESE), and 95 per cent confidence interval coverage percent (95 per cent CI Cov.).

						β_1 (continuous)				β_2 (binary)			
n	**True $\beta^{(0)}$**	**True $\rho_{sp}^{(r)}$**	**Assumed $\beta^{(0)}$**	**Correct Assumption?**	**Method**	**Bias**	**SE**	**ESE**	**95 per cent CI Cov.**	**Bias**	**SE**	**ESE**	**95 per cent CI Cov.**
100	$\beta^{(1)}$	0.5	$\beta^{(1)}$	Yes	ML-Model 1	<1	0.127	0.127	94.6	<1	0.273	0.274	94.5
			$\beta^{(1)}$	Yes	MI-Model 1	<1	0.129	0.130	95.0	1	0.276	0.283	94.4
			$\alpha^{(0)}+\lambda_2$	No*	ML-Model V(2)	−39	0.175	0.175	37.8	20	0.351	0.335	92.6
			$\alpha^{(0)}+\lambda_2$	No*	MI-Model V(2)	−37	0.178	0.165	49.2	18	0.349	0.344	92.9
			$\alpha^{(0)}+\lambda_2$	No†	ML-Model V(2)	34	0.175	0.168	50.8	−38	0.360	0.338	84.5
			$\alpha^{(0)}+\lambda_2$	No†	MI-Model V(2)	34	0.177	0.182	56.4	−35	0.365	0.345	86.8
			$(\alpha^{(0)}-\lambda_3)/\gamma$	No*	ML-Model V(3)	−56	0.440	0.250	81.3	27	0.701	0.447	97.9
			$(\alpha^{(0)}-\lambda_3)/\gamma$	No*	MI-Model V(3)	−66	0.370	0.316	66.0	30	0.580	0.542	95.8
			$(\alpha^{(0)}-\lambda_3)/\gamma$	No†	ML-Model V(3)	50	0.447	0.246	88.1	−55	0.808	0.474	98.3
			$(\alpha^{(0)}-\lambda_3)/\gamma$	No†	MI-Model V(3)	58	0.361	0.323	71.6	−63	0.656	0.586	92.2
100	$\beta^{(1)}$	0.8	$\beta^{(1)}$	Yes	ML-Model 1	<1	0.126	0.135	92.2	<1	0.274	0.281	94.2
			$\beta^{(1)}$	Yes	MI-Model 1	<1	0.120	0.126	94.3	−1	0.262	0.269	94.0
			$\alpha^{(0)}+\lambda_2$	No*	ML-Model V(2)	−38	0.154	0.151	28.3	19	0.312	0.304	91.2
			$\alpha^{(0)}+\lambda_2$	No*	MI-Model V(2)	−30	0.188	0.164	64.0	15	0.305	0.289	93.1
			$\alpha^{(0)}+\lambda_2$	No†	ML-Model V(2)	35	0.157	0.157	40.3	−35	0.327	0.318	82.4
			$\alpha^{(0)}+\lambda_2$	No†	MI-Model V(2)	26	0.181	0.163	70.0	−28	0.328	0.306	89.3
			$(\alpha^{(0)}-\lambda_3)/\gamma$	No*	ML-Model V(3)	−41	0.225	0.185	52.0	20	0.387	0.328	95.2
			$(\alpha^{(0)}-\lambda_3)/\gamma$	No*	MI-Model V(3)	−33	0.207	0.190	66.1	16	0.325	0.304	93.7
			$(\alpha^{(0)}-\lambda_3)/\gamma$	No†	ML-Model V(3)	38	0.227	0.189	62.4	−38	0.422	0.358	91.0
			$(\alpha^{(0)}-\lambda_3)/\gamma$	No†	MI-Model V(3)	29	0.200	0.190	71.6	−31	0.355	0.330	91.2
100	$\alpha^{(0)}$	0.5	$\beta^{(1)}$	No	ML-Model 1	63	0.127	0.127	13.8	−16	0.272	0.285	87.7
			$\beta^{(1)}$	No	MI-Model 1	65	0.127	0.136	15.9	−16	0.274	0.278	89.1
			$\alpha^{(0)}+\lambda_2$	Yes*	ML-Model V(2)	<1	0.173	0.172	93.9	<1	0.348	0.356	93.5

n	True $\beta^{(0)}$	True $\rho_{sp}^{(r)}$	Assumed $\beta^{(0)}$	Correct Assumption?	Method	β_1 (continuous)				β_2 (binary)			
						Bias	SE	ESE	95 per cent CI Cov.	Bias	SE	ESE	95 per cent CI Cov.
			$\alpha^{(0)}+\lambda_2$	Yes*	MI-Model V(2)	4	0.177	0.169	94.3	<1	0.348	0.334	96.8
			$(\alpha^{(0)}-\lambda_3)/\gamma$	Yes*	ML-Model V(3)	−30	0.442	0.265	99.5	7	0.685	0.475	98.2
			$(\alpha^{(0)}-\lambda_3)/\gamma$	Yes*	MI-Model V(3)	−40	0.378	0.333	96.0	11	0.598	0.531	97.4
100	$\alpha^{(0)}$	0.8	$\beta^{(1)}$	No	ML-Model 1	63	0.126	0.127	14.9	−14	0.273	0.273	90.3
			$\beta^{(1)}$	No	MI-Model 1	63	0.120	0.132	15.9	−16	0.261	0.286	86.6
			$\alpha^{(0)}+\lambda_2$	Yes*	ML-Model V(2)	<1	0.154	0.149	95.5	1	0.313	0.318	94.2
			$\alpha^{(0)}+\lambda_2$	Yes*	MI-Model V(2)	14	0.189	0.168	93.0	−4	0.303	0.285	94.9
			$(\alpha^{(0)}-\lambda_3)/\gamma$	Yes*	ML-Model V(3)	−6	0.231	0.192	96.8	3	0.389	0.348	97.5
			$(\alpha^{(0)}-\lambda_3)/\gamma$	Yes*	MI-Model V(3)	8	0.208	0.193	93.9	−2	0.322	0.302	95.7
100	$\alpha^{(0)}+0.75$	0.5	$\beta^{(1)}$	No	ML-Model 1	−25	0.126	0.130	23.2	55	0.271	0.284	71.7
			$\beta^{(1)}$	No	MI-Model 1	−25	0.127	0.123	27.9	58	0.275	0.284	72.5
			$\alpha^{(0)}+\lambda_2$	Yes†	ML-Model V(2)	<1	0.174	0.171	95.0	−1	0.358	0.370	93.7
			$\alpha^{(0)}+\lambda_2$	Yes†	MI-Model V(2)	<1	0.178	0.170	95.1	2	0.366	0.355	95.4
			$(\alpha^{(0)}-\lambda_3)/\gamma$	Yes†	ML-Model V(3)	12	0.439	0.252	99.8	−31	0.817	0.517	98.2
			$(\alpha^{(0)}-\lambda_3)/\gamma$	Yes†	MI-Model V(3)	17	0.353	0.312	97.6	−38	0.655	0.626	97.5
100	$\alpha^{(0)}+0.75$	0.8	$\beta^{(1)}$	No	ML-Model 1	−25	0.126	0.125	23.7	57	0.272	0.293	72.0
			$\beta^{(1)}$	No	MI-Model 1	−26	0.120	0.133	22.4	58	0.261	0.285	70.4
			$\alpha^{(0)}+\lambda_2$	Yes†	ML-Model V(2)	<1	0.156	0.152	95.0	<1	0.325	0.328	94.2
			$\alpha^{(0)}+\lambda_2$	Yes†	MI-Model V(2)	−5	0.168	0.164	93.6	11	0.329	0.306	95.1
			$(\alpha^{(0)}-\lambda_3)/\gamma$	Yes†	ML-Model V(3)	2	0.227	0.183	97.3	−4	0.418	0.365	97.1
			$(\alpha^{(0)}-\lambda_3)/\gamma$	Yes†	MI-Model V(3)	−3	0.200	0.191	93.6	6	0.356	0.333	95.6
250	$\beta^{(1)}$	0.5	$\beta^{(1)}$	Yes	ML-Model 1	<1	0.079	0.083	92.5	<1	0.172	0.173	95.3
			$\beta^{(1)}$	Yes	MI-Model 1	<1	0.080	0.080	95.1	<1	0.174	0.179	94.9
			$\alpha^{(0)}+\lambda_2$	No*	ML-Model V(2)	−39	0.105	0.106	4.1	19	0.213	0.212	86.5
			$\alpha^{(0)}+\lambda_2$	No*	MI-Model V(2)	−39	0.102	0.106	5.7	19	0.215	0.211	86.9
			$\alpha^{(0)}+\lambda_2$	No†	ML-Model V(2)	34	0.105	0.102	9.2	−36	0.221	0.225	62.7

n	True $\beta^{(0)}$	True $\rho_{sp}^{(r)}$	Assumed $\beta^{(0)}$	Correct Assumption?	Method	β_1 (continuous)				β_2 (binary)			
						Bias	SE	ESE	95 per cent CI Cov.	Bias	SE	ESE	95 per cent CI Cov.
			$\alpha^{(0)}+\lambda_2$	No†	MI-Model V(2)	34	0.104	0.107	9.9	−35	0.220	0.217	65.9
			$(\alpha^{(0)}-\lambda_3)/\gamma$	No*	ML-Model V(3)	−66	0.257	0.206	13.1	32	0.357	0.308	91.4
			$(\alpha^{(0)}-\lambda_3)/\gamma$	No*	MI-Model V(3)	−68	0.216	0.228	10.1	33	0.339	0.336	87.9
			$(\alpha^{(0)}-\lambda_3)/\gamma$	No†	ML-Model V(3)	60	0.247	0.199	18.8	−61	0.408	0.356	71.7
			$(\alpha^{(0)}-\lambda_3)/\gamma$	No†	MI-Model V(3)	60	0.210	0.220	15.6	−61	0.369	0.362	67.0
250	$\beta^{(1)}$	0.8	$\beta^{(1)}$	Yes	ML-Model 1	<1	0.079	0.080	93.8	−1	0.171	0.173	94.4
			$\beta^{(1)}$	Yes	MI-Model 1	<1	0.078	0.083	94.6	<1	0.169	0.180	93.2
			$\alpha^{(0)}+\lambda_2$	No*	ML-Model V(2)	−39	0.093	0.095	1.5	18	0.190	0.185	84.3
			$\alpha^{(0)}+\lambda_2$	No*	MI-Model V(2)	−35	0.124	0.107	29.7	17	0.190	0.176	88.5
			$\alpha^{(0)}+\lambda_2$	No†	ML-Model V(2)	34	0.093	0.092	3.2	−35	0.198	0.194	56.3
			$\alpha^{(0)}+\lambda_2$	No†	MI-Model V(2)	31	0.120	0.102	31.1	−32	0.215	0.200	67.6
			$(\alpha^{(0)}-\lambda_3)/\gamma$	No*	ML-Model V(3)	−40	0.115	0.113	2.6	19	0.204	0.192	87.4
			$(\alpha^{(0)}-\lambda_3)/\gamma$	No*	MI-Model V(3)	−36	0.131	0.119	30.3	17	0.195	0.181	88.3
			$(\alpha^{(0)}-\lambda_3)/\gamma$	No†	ML-Model V(3)	36	0.110	0.107	5.4	−36	0.217	0.208	60.3
			$(\alpha^{(0)}-\lambda_3)/\gamma$	No†	MI-Model V(3)	32	0.126	0.112	31.1	−34	0.222	0.209	68.4
250	$\alpha^{(0)}$	0.5	$\beta^{(1)}$	No	ML-Model 1	63	0.079	0.082	0.7	−16	0.172	0.170	81.5
			$\beta^{(1)}$	No	MI-Model 1	63	0.080	0.081	0.7	−16	0.174	0.185	80.4
			$\alpha^{(0)}+\lambda_2$	Yes*	ML-Model V(2)	<1	0.106	0.108	94.7	<1	0.214	0.212	95.6
			$\alpha^{(0)}+\lambda_2$	Yes*	MI-Model V(2)	<1	0.101	0.103	93.5	<1	0.214	0.219	94.7
			$(\alpha^{(0)}-\lambda_3)/\gamma$	Yes*	ML-Model V(3)	−44	0.257	0.213	95.1	11	0.363	0.325	98.1
			$(\alpha^{(0)}-\lambda_3)/\gamma$	Yes*	MI-Model V(3)	−47	0.220	0.218	88.1	11	0.335	0.331	95.0
250	$\alpha^{(0)}$	0.8	$\beta^{(1)}$	No	ML-Model 1	63	0.079	0.078	0.1	−16	0.171	0.172	79.2
			$\beta^{(1)}$	No	MI-Model 1	63	0.078	0.082	0.5	−16	0.170	0.178	79.9
			$\alpha^{(0)}+\lambda_2$	Yes*	ML-Model V(2)	<1	0.093	0.087	96.0	<1	0.191	0.188	94.5
			$\alpha^{(0)}+\lambda_2$	Yes*	MI-Model V(2)	6	0.123	0.107	95.9	−1	0.191	0.185	94.9
			$(\alpha^{(0)}-\lambda_3)/\gamma$	Yes*	ML-Model V(3)	−3	0.115	0.105	97.7	<1	0.206	0.197	96.0

n	True $\beta^{(0)}$	True $\rho_{sp}^{(r)}$	Assumed $\beta^{(0)}$	Correct Assumption?	Method	β_1 (continuous) Bias	SE	ESE	95 per cent CI Cov.	β_2 (binary) Bias	SE	ESE	95 per cent CI Cov.
			$(\alpha^{(0)}-\lambda_3)/\gamma$	Yes*	MI-Model V(3)	3	0.129	0.121	95.6	<1	0.196	0.191	94.5
250	$\alpha^{(0)}+0.75$	0.5	$\beta^{(1)}$	No	ML-Model 1	−25	0.079	0.080	1.6	57	0.170	0.175	44.2
			$\beta^{(1)}$	No	MI-Model 1	−26	0.081	0.080	2.3	56	0.174	0.179	48.7
			$\alpha^{(0)}+\lambda_2$	Yes†	ML-Model V(2)	<1	0.105	0.104	95.3	1	0.220	0.213	94.4
			$\alpha^{(0)}+\lambda_2$	Yes†	MI-Model V(2)	<1	0.104	0.106	94.9	<1	0.220	0.222	95.5
			$(\alpha^{(0)}-\lambda_3)/\gamma$	Yes†	ML-Model V(3)	18	0.245	0.209	95.4	−38	0.405	0.338	97.2
			$(\alpha^{(0)}-\lambda_3)/\gamma$	Yes†	MI-Model V(3)	20	0.216	0.227	86.8	−44	0.374	0.377	93.0
250	$\alpha^{(0)}+0.75$	0.8	$\beta^{(1)}$	No	ML-Model 1	−26	0.079	0.080	1.2	58	0.171	0.175	41.1
			$\beta^{(1)}$	No	MI-Model 1	−25	0.077	0.080	1.3	55	0.170	0.174	48.3
			$\alpha^{(0)}+\lambda_2$	Yes†	ML-Model V(2)	<1	0.093	0.091	95.1	<1	0.198	0.198	94.8
			$\alpha^{(0)}+\lambda_2$	Yes†	MI-Model V(2)	−2	0.118	0.100	95.5	4	0.210	0.195	95.6
			$(\alpha^{(0)}-\lambda_3)/\gamma$	Yes†	ML-Model V(3)	1	0.112	0.110	96.0	−2	0.219	0.212	95.4
			$(\alpha^{(0)}-\lambda_3)/\gamma$	Yes†	MI-Model V(3)	−1	0.123	0.116	95.0	2	0.218	0.212	95.1

ML=maximum likelihood, MI=multiple imputation. $\lambda_2=\beta^{(1)}-\alpha^{(1)}$, $\lambda_3=\alpha^{(1)}-\gamma\beta^{(1)}$.

* $\alpha^{(1)}$ specified so that $\alpha^{(0)}+\lambda_2=(\alpha^{(0)}-\lambda_3)/\gamma=\alpha^{(0)}$.

† $\alpha^{(1)}$ specified so that $\alpha^{(0)}+\lambda_2=(\alpha^{(0)}-\lambda_3)/\gamma=\alpha^{(0)}+0.75$.

Table III

BHS-2 study results (n=248). IADL dependency (range: 0–7) regressed on patient sex (1=male,0=female) and age (years); estimated coefficients ($\hat{\beta}$) and standard errors (SE).

Validation Data Used?	Assumed $\beta^{(0)}$	Method	λ_2	$\rho^{(0)}_{(sp)}$	Sex		Age (Years)	
					$\hat{\beta}$	SE	$\hat{\beta}$	SE
No	$\beta^{(1)}$	Subject Only	—	—	1.14	0.52	0.10	0.02
	$\alpha^{(0)}$	Subject + Proxy	—	—	0.81	0.37	0.12	0.02
	$\beta^{(1)}$	ML-Model 1	—	—	1.14	0.52	0.10	0.02
	$\beta^{(1)}$	MI-Model 1	—	0.35	1.08	0.51	0.10	0.02
	$\beta^{(1)}$	MI-Model 1	—	0.70	1.11	0.54	0.10	0.02
	$\alpha^{(0)}$	ML-Model 2	(0, 0)	—	0.81	0.38	0.12	0.02
	$\alpha^{(0)}$	MI-Model 2	(0, 0)	0.35	0.80	0.45	0.12	0.02
	$\alpha^{(0)}$	MI-Model 2	(0, 0)	0.70	0.83	0.41	0.12	0.02
	$\alpha^{(0)}+\lambda_2$	ML-Model 2	(−0.26,0.01)	—	0.81	0.38	0.13	0.02
	$\alpha^{(0)}+\lambda_2$	MI-Model 2	(−0.26,0.01)	0.35	0.82	0.47	0.13	0.02
	$\alpha^{(0)}+\lambda_2$	MI-Model 2	(−0.26,0.01)	0.70	0.80	0.43	0.13	0.02
	$\alpha^{(0)}+\lambda_2$	ML-Model 2	(−0.52,0.02)	—	0.82	0.38	0.14	0.02
	$\alpha^{(0)}+\lambda_2$	MI-Model 2	(−0.52,0.02)	0.35	0.82	0.49	0.14	0.02
	$\alpha^{(0)}+\lambda_2$	MI-Model 2	(−0.52,0.02)	0.70	0.84	0.45	0.14	0.02
	$\alpha^{(0)}$	ML-Model 3	(0, 0)	0.35	0.81	0.86	0.12	0.05
	$\alpha^{(0)}$	MI-Model 3	(0, 0)	0.35	0.83	0.80	0.12	0.04
	$\alpha^{(0)}$	ML-Model 3	(0, 0)	0.70	0.81	0.49	0.12	0.03
	$\alpha^{(0)}$	MI-Model 3	(0, 0)	0.70	0.80	0.47	0.12	0.02
	$\alpha^{(0)}/\gamma$	ML-Model 3	(0, 0)	0.35	3.34	1.38	0.33	0.10
	$\alpha^{(0)}/\gamma$	MI-Model 3	(0, 0)	0.35	3.26	1.14	0.33	0.06
	$\alpha^{(0)}/\gamma$	ML-Model 3	(0, 0)	0.70	1.41	0.55	0.17	0.03
	$\alpha^{(0)}/\gamma$	MI-Model 3	(0, 0)	0.70	1.40	0.55	0.17	0.03
	$(\alpha^{(0)}-\lambda_3)/\gamma$	ML-Model 3	(−0.26,0.01)	0.35	0.81	0.86	0.13	0.05
	$(\alpha^{(0)}-\lambda_3)/\gamma$	MI-Model 3	(−0.26,0.01)	0.35	0.90	0.83	0.13	0.04

Validation Data Used?	Assumed $\beta^{(0)}$	Method	λ_2	$\rho^{(0)}_{(sp)}$	Sex		Age (Years)	
					$\hat{\beta}$	SE	$\hat{\beta}$	SE
	$(\alpha^{(0)}-\lambda_3)/\gamma$	ML-Model 3	(−0.26,0.01)	0.70	0.81	0.50	0.13	0.03
	$(\alpha^{(0)}-\lambda_3)/\gamma$	MI-Model 3	(−0.26,0.01)	0.70	0.84	0.49	0.13	0.02
	$(\alpha^{(0)}-\lambda_3)/\gamma$	ML-Model 3	(−0.52,0.02)	0.35	0.82	0.86	0.14	0.05
	$(\alpha^{(0)}-\lambda_3)/\gamma$	MI-Model 3	(−0.52,0.02)	0.35	0.76	0.79	0.14	0.04
	$(\alpha^{(0)}-\lambda_3)/\gamma$	ML-Model 3	(−0.52,0.02)	0.70	0.82	0.50	0.14	0.03
	$(\alpha^{(0)}-\lambda_3)/\gamma$	MI-Model 3	(−0.52,0.02)	0.70	0.85	0.49	0.14	0.03
Yes	$\alpha^{(0)}+\lambda_2$	ML-Model V(2)	$\beta_1-\hat{\alpha}_1$	$\widehat{\rho}^{(1)}_{(sp)}$	1.14	0.47	0.14	0.02
	$\alpha^{(0)}+\lambda_2$	MI-Model V(2)	$\beta_1-\hat{\alpha}_1$	$\widehat{\rho}^{(1)}_{(sp)}$	1.18	0.53	0.14	0.03
	$(\alpha^{(0)}-\lambda_3)/\gamma$	ML-Model V(3)	$\beta_1-\hat{\alpha}_1$	$\widehat{\rho}^{(1)}_{(sp)}$	1.14	0.50	0.14	0.02
	$(\alpha^{(0)}-\lambda_3)/\gamma$	MI-Model V(3)	$\beta_1-\hat{\alpha}_1$	$\widehat{\rho}^{(1)}_{(sp)}$	1.14	0.56	0.14	0.03

'Subject Only' refers to linear regression with only observed subject data, 'Subject+Proxy' refers to linear regression where proxy data substitute for missing subject data, ML=maximum likelihood, MI=multiple imputation. Models 2 and 3 without validation data: assume $\rho^{(0)}_{(sp)}=0.35$ or 0.70, λ_2=(−0.26,0.01) or (−0.52,0.02). $\rho^{(0)}_{(sp)}$ and λ_2 used to derive λ_3. Models V(2) and V(3) with validation data: estimate $\rho^{(0)}_{(sp)}$ as $\widehat{\rho}^{(1)}_{(sp)}=0.87$, estimate λ_2 as $\hat{\beta}^{(1)}-\hat{\alpha}^{(1)}$=(0.52,0.04), and use $\lambda_2$to derive λ_3.