5-2016

Multiple Imputation When Rate of Change Is The Outcome of Interest

Manisha Desai
Stanford University, manishad@stanford.edu

Aya A. Mitani
Boston University, aya.mitani@gmail.com

Susan W. Bryson
Stanford University, swbryson@stanford.edu

Thomas Robinson
Stanford University, tom.robinson@stanford.edu

Multiple Imputation When Rate of Change Is The Outcome of Interest

Multiple Imputation When Rate of Change is the Outcome of Interest

Manisha Desai
Stanford University
Stanford, CA

Aya A. Mitani
Boston University
Boston, MA

Susan W. Bryson
Stanford University
Stanford, CA

Thomas Robinson
Stanford University
Stanford, CA

Little research has been devoted to multiple imputation (MI) of derived variables. This study investigates various MI approaches for the outcome, rate of change, when the analysis model is a two-stage linear regression. Simulations showed that competitive approaches depended on the missing data mechanism and presence of auxiliary terms.

Keywords: Multiple imputation, derived variables, active imputation, passive imputation, mixed effects model, two-stage linear regression, slope

Introduction

Multiple imputation (MI) techniques, although sound in theory, often present challenges in practice. One issue involves how best to multiply impute derived variables. These include interaction terms, higher-order terms, variables that are functions of others such as body mass index (BMI) – a function of height and weight – and rate of change in a measure (or the slope). Consider interaction terms. A natural question is whether the main effects should first be imputed so that the interaction term can then be derived, or whether the interaction terms should be imputed as any other variable. The former approach falls under the umbrella of what is called passive imputation methods and the latter under the umbrella of active imputation techniques (von Hippel, 2009; van Buuren, 2012). Similar issues arise in other contexts. For example, if one has data on height but not on weight for a subject, where BMI is of interest, a passive approach would be to first impute weight and then derive BMI, while an active approach would impute BMI directly as any other variable along with weight.

This work is motivated by the ongoing Stanford GOALS randomized clinical trial, which is designed to evaluate a novel intervention in reducing BMI

(weight in kilograms divided by the square of height in meters) among overweight and obese children (Robinson et al., 2013). The primary outcome of the trial is the rate of change in BMI, which entails measuring BMI at 4 time points over a 3-year period (at baseline and at approximately 1, 2, and 3 years post-randomization). As typically occurs in clinical trials, it is anticipated that a proportion of children will be missing one or more follow-up BMI measurements upon which to calculate the rate of change. The focus of this study is on how to impute slopes when slope is the outcome of interest and missing for a proportion of subjects due to incomplete BMI measurements.

Multiple imputation (MI)

MI is a simulation-based method for handling missing data (Rubin, 1996). There are three main steps involved in its use. The first is the imputation step where plausible values for missing data are drawn and filled in to create a full data set. This is done multiple, or m, times. The second step is the model-fitting step where the scientific model is fitted to each of the m data sets. The third step is the combination step where Rubin's rules are applied to provide one summary result (Little & Rubin, 2002). Standard MI produces results that are statistically valid when the data are missing at random (MAR) (Ibid.). Essentially this means that missingness can be related to observed features, although conditional on such features, missingness is not related to unobserved data. The data are considered missing completely at random (MCAR) if missingness is unrelated to both unobserved and observed features, and the data are not missing at random (NMAR) if missingness is related to unobserved features. See Rubin's seminal text on missing data for a complete description on mechanisms of missingness (Ibid.). MI may also be applied when the data are NMAR, but this requires explicit modeling of the missing data mechanism (Ibid.). Further, even when the data are NMAR, in the presence of sufficiently strong auxiliary data, applying MI assuming MAR may be reasonable (Collins et al., 2001). Importantly, MI yields estimates with desirable statistical properties under the same conditions in which the more ideal maximum likelihood-based methods for handling missing data produce estimates (Little & Rubin, 2002). MI, in contrast, is more accessible as it is available in several mainstream packages, while software is not available for all cases when applying maximum likelihood-based approaches.

Specification of the imputation model (Step 1 of the MI process described above) is a critical step when applying MI, and requires the user to make key decisions. This includes the choice of one of two main approaches for imputing

the data. These are the joint modeling (JM) approach and the fully conditional specification (FCS) approach (van Buuren, 2007). Briefly, the JM approach involves specification of a joint model for the data from which the posterior predictive distribution for the missing data can be derived in order to impute the missing data. The approach typically assumes a multivariate normal distribution (although other parametric distributions are possible) and has well established and tractable statistical properties (Ibid.). FCS was developed to accommodate data sets of mixed type (i.e., when categorical variables are present), and therefore is useful when a plausible multivariate distribution may not be appropriate. Although the statistical properties are not tractable, use of FCS has been shown to be comparable to that of JM in practice and through simulation studies (Ibid.). In his excellent text on MI, van Buuren (2012) described other considerations, including which variables to include as predictors, the number of imputations to be performed, and how to handle derived variables.

Current literature on active versus passive imputation approaches

There are two main umbrellas of approaches that can be used to impute derived variables: active and passive. In general, active imputation involves imputing the derived variable as just another variable or JAV method (Seaman et al., 2012). In contrast, purely passive imputation – as implemented in STATA through ICE, which was developed by Patrick Royston (and more recently through MI IMPUTE CHAIN) or MICE in R – is an iterative process that involves using the derived variable in the imputation of all variables with the exception of the main effects (van Buuren, 2012; Royston, 2009). The derived term is then created once the main effects have been imputed (Ibid.). The advantage of an active approach is that it adheres to the generally well-accepted rule that the relationships specified in the scientific model are included in the imputation model (i.e., that the imputation model is congenial with the scientific model leading to what is called proper imputations (van Buuren, 2012; Rubin, 1996; Rubin, 1987; Meng, 1994). For a more detailed discussion on this topic see Rubin (1987) and Meng (1994). This ensures that interrelationships are preserved appropriately, leading to unbiased descriptions of the relationship of interest. Passive approaches, on the other hand, ensure that imputed values adhere to consistent relationships and are therefore plausible. For example, under active imputation, if BMI is imputed as just another variable, it is possible to get a value of say 30 instead of 20 when observed weight is 54.5 kg (120 lbs.) and imputed height is 1.65 meters (65 inches). Passive imputation can prevent such inconsistencies.

Research on optimal approaches for imputing derived variables has been limited. Much of the literature has focused on one of the more common derived variables – interaction terms (von Hippel, 2009; Seaman et al., 2012; Allison, 2002). Allison (2002) briefly introduced the problem of multiply imputing interaction terms in his widely used and practical textbook on missing data. von Hippel (2009) discussed two major approaches for handling missing data of interaction and higher-order terms under an assumption that the data are MCAR – a "transform then impute" or active approach and an "impute then transform" or passive-like approach, and provided a mathematical argument to demonstrate that under MCAR the active approach is less biased than the passive. von Hippel (2009) also argued that the purely passive approach is similarly biased. White at al. (2011) compared passive and active approaches for interaction terms under MAR and through simulations demonstrated that they are both biased but comparable. The authors recommend sensitivity analyses of both approaches in assessing robustness of findings. Mitani and others (2015) evaluated active and passive methods across different software platforms for interaction terms that involved categorical variables. They found an improved passive approach – where an interaction between the outcome and main effect is included in the imputation of the other main effect is included in a passive imputation -- outperformed both active and purely passive approaches. van Buuren (2012) emphasizes the importance of internally consistent imputed values, and therefore recommends passive approaches that are designed for this purpose. Although active approaches may yield valid statistical inferences under certain conditions, he argues that if implausible combinations of imputed values are produced that would not have been observed otherwise it undermines the imputation process. van Buuren (2012) discussed active versus passive imputation in the context of interaction terms, sum scores, and other derived variables such as ratios. However, their variability in performance for measures of rate of change or slope was not discussed. Although imputation in the context of mixed effects models has received some attention (e.g., Schafer, 1997b), no research has been conducted on imputation when the scientific model is a two-stage linear regression. Our paper focuses on situations in which slope – a variable derived from repeated measurements – is the outcome of interest and the scientific model is a two-stage linear regression. Furthermore we consider a wider range of active and passive methods under a variety of missing data mechanisms that include the NMAR condition.

The Stanford GOALS trial

This work is motivated by a large-scale community-based randomized clinical trial currently being conducted at Stanford called Stanford GOALS (Robinson et al., 2013). The objective of the trial is to evaluate a novel multi-level, multi-component intervention to treat overweight and obese children aged 7-11 years old. The intervention consists of access to after-school team sports, home visits for behavioral counseling and modifications of the home environment, and primary care counseling. Two-hundred forty subjects are randomized to 1 of 2 groups – the intervention group or a control group receiving health and nutrition education plus their current standard of care. The primary outcome is a derived measure – the rate of change in BMI. Baseline BMI is measured for all randomized subjects. BMI is then measured at approximately 1, 2, and 3 years post randomization, and the outcome is a function of the 4 BMI measurements. As with most prospective randomized clinical trials, some subjects are expected to be missing one or more of the follow-up BMI measures.

The data are not expected to be missing completely at random. Missingness may be related to follow-up time, treatment arm, baseline BMI and even rate of change in BMI itself. However, conditional on baseline and intermittent BMI values, as well as a number of related measures including triceps skinfold, waist circumference, blood pressure, fasting blood lipids, insulin, glucose, and hemoglobin A1c to name a few, recorded at baseline and annually, we expect the data to be MAR. Our analytic plan specifies that we will use MI to impute the outcome or rate of change. For subjects who have at least 2 measurements, slope can be calculated with the baseline and subsequent measurements. Thus, the plan is to impute slope for those subjects who only have a baseline BMI and no other follow-up measurements.

Slope model versus mixed effects models

A simplified version of the primary analysis proposed for Stanford GOALS is a two-stage linear model or what we will refer to here as a slope model. More specifically, in the first stage, we derive a slope for all individuals with at least two BMI measurements by regressing BMI on time for each individual:

$$BMI_{ij} = \alpha_{0i} + \alpha_{1i}\,time_{ij} + \varepsilon_{ij}, \tag{1a}$$

where $time_{ij}$ represents time in months since randomization for the ith person at the jth visit, and ε_{ij} represents normally distributed random error; and in the second, we regress slope on treatment arm:

$$\hat{\alpha}_{1i} = \eta_0 + \eta_1 Treatment_i + \delta_i, \qquad (1b)$$

where $Treatment_i$ is an indicator for whether the ith subject is in the intervention group, and δ_i represents normally distributed random error. As a secondary analysis, we will fit a mixed effects model where we regress BMI on treatment arm, time, and an interaction between time and treatment, and include subject-specific random effects for the intercept and slope with time:

$$BMI_{ij} = (\gamma_{0i} + \beta_0) + \beta_1 Treatment_i + (\beta_2 + \gamma_{1i})time_{ij} + \beta_3 Treatment_i \times time_{ij} + \xi_{ij}, (2)$$

where γ_{0i} and γ_{1i} are normally-distributed subject-specific random effects – a random intercept and slope for time for the ith subject – and ξ_{ij} is the normally distributed error term for the jth observation corresponding to the ith subject. The parameter of interest is the difference in rate of change in BMI between the treated and non-treated groups. For the slope model, this is η_1 and for the mixed effects model, this is β_3. Although equivalent in that both models assume subjects have their own random intercept and random slope, the procedures for fitting these models differ and may yield different estimates particularly in the presence of missing data. The mixed effects model with random intercepts and random slopes or RIRS model – yields an estimate based on any subject who contributes at least one outcome of BMI and relies on an assumption that the data are missing at random. The slope model yields an estimate based on any subject who contributes at least two outcomes of BMI (at baseline and at least one follow-up time point). Thus, those who contribute only a baseline measurement are excluded from the analysis. Incorporating MI methods into the process to impute rate of change for those missing all 3 follow-up BMI values and then applying the slope model relies on a MAR assumption about the missing data as well.

Issues with choosing a two-stage linear regression or slope model over the RIRS model have been discussed at great length (e.g., Rogosa & Saner, 1995; Gelman, 2005; Jusko & Shively, 2013; Duch & Stevenson, 2005). There are several reasons, however, why the slope model might be chosen over the RIRS model (Rogosa & Saner, 1995; Gelman, 2005; Liao et al., 2013). Specialized software is required in the latter case and the complexity of the model parameterization can compromise interpretability; in contrast, Equation 1b above

is easy to interpret compared to a mixed effects model that in a longitudinal setting involves interpreting an interaction term. In addition, in the process of fitting the slope model, insight into the data can be gained; in particular, heterogeneity in estimates of intercepts and slopes across subjects can be assessed. These features are not typically examined when fitting the RIRS model; rather, they are treated as nuisance parameters. There may be advantages, however, in the properties of estimates resulting from RIRS models over those resulting from the two-stage models. Gelman (2005) advocated the two-stage model over the mixed effects model in certain conditions, because of the simplicity and insight gained which are strong reasons to choose the two-stage model when it is appropriate. It was noted there are situations where the advantages of the mixed effects model outweigh its complexity – e.g., in the longitudinal setting, if data within individuals are sufficiently sparse, it may be desirable to borrow strength from other individuals by fitting the RIRS model.

Therefore, the goal is to determine whether different approaches to imputing slope – when a two-stage linear model or slope model is of interest – affect properties of estimates of interest. In particular we are interested in approaches that can be employed by mainstream packages that have incorporated MI procedures. To that end, variability will be evaluated among active versus passive methods of imputation that are readily accessible in STATA and R. We focus on situations where the treatment effect is of interest and where auxiliary information may or may not be present.

Methodology

A simulation study was conducted to closely mimic the Stanford GOALS trial, where there were two treatment arms with 120 subjects per group. Other parameters were not based on empirical findings, as the study is currently ongoing. Specifically, we assumed there was a treatment effect such that those in the treatment group did not increase their BMI over time, while those in the control group increased by 0.5 BMI units per year, the percent of observations (not individuals) missing varied (20% and 40%), the true model varied (RI or slope/RIRS model), and auxiliary information was and was not present (Table 1).

Table 1. Description of Scenarios

Scenario	Model for Data Generation	Model for Missingness	Average % Missing	Auxiliary Data Available	Auxiliary Data Related to Variables	Auxiliary Data Related to Missingness
A1	RI model	MAR	40	No	N/A	N/A
A2	RI model	NMAR	40	No	N/A	N/A
B3	RI model	MAR	40	Yes	Yes	No
B4	RI model	MAR	40	Yes	Yes	Yes
B5	RI model	NMAR	40	Yes	Yes	No
B6	RI model	NMAR	40	Yes	Yes	Yes
B7	RI model	NMAR	20	Yes	Yes	Yes
C8	RIRS/Slope model	MAR	40	Yes	Yes	Yes
C9	RIRS/Slope model	NMAR	40	Yes	Yes	Yes

Generating BMI

The true model from which the data were generated under Scenarios A and B was a mixed effects model with random intercept only – which we will refer to as the RI model – a special case of the RIRS model which includes both random intercept and random slope for time for each subject (Table 1). The RI model can be expressed as:

$$BMI_{ij} = (\gamma_{0i} + \beta_0) + \beta_1 Treatment_i + \beta_2 time_{ij} + \beta_3 Treatment_i \times time_{ij} + \xi_{ij},$$

where $\beta_3 = -0.5$, $\xi_{ij} \sim N(0,1)$, $\gamma_{0i} \sim N(0,2)$.

The true model from which the data were generated under Scenario C was the slope model, so that we first generated slopes from the following model:

$$\alpha_{1i} = \eta_0 + \eta_1 Treatment_i + \delta_i$$

where $\eta_1 = -0.5$, $\delta_i \sim N(0,0.7)$. BMI values for each subject were than generated based on the realized slope values and random intercepts according to the following model:

$$BMI_{ij} = \alpha_{0i} + \hat{\alpha}_{1i} time_{ij} + \varepsilon_{ij}$$

where $\alpha_{0i} \sim N(25,2)$, $\varepsilon_{ij} \sim N(0,1)$, and $time_{ij}$ takes on values of j=0, 1, 2, or 3 to represent the jth visit of the ith individual. Note that this is also the RIRS model, as

each individual has its own intercept and slope generated as random variables from a normal distribution.

Generating auxiliary terms

Two auxiliary terms – one continuous to represent age, and one dichotomous to represent gender – were generated. In some scenarios (B3 and B5), the auxiliary variables were related to baseline BMI and the rate of change in BMI or the slope, and in others they were additionally related to missingness (B4, B6, B7, C8, C9) (Table 1). Specifically, in Scenarios B3-B7 and C8-C9, one continuous variable representing Age, was generated for the ith person as follows:

$$Age_i = 5 \times Slope_i + 2 \times BaselineBMI_i + \varepsilon_{ij}$$

where the error was normally distributed with variance of 4. An indicator for male gender was generated as follows:

$$logit(Male_i) = -5 + 0.5 \times Slope_i + 0.2 \times BaselineBMI_i$$

Inducing missingness

Missingness (under MAR) was related to time since randomization and treatment arm for Scenarios A1 and B3:

$$logit(missing_{ij}|Treatment_i = control) = -1.5 + 2 \times time_{ij}, \text{ and}$$
$$logit(missing_{ij}|Treatment_i = intervention) = 2.$$

Missingness (under MAR) was additionally related to auxiliary terms for Scenarios B4 and C8:

$$logit(missing_{ij}|Treatment_i = control) = -3 + 2 \times time_{ij} + 1.5 \times I[Age_i < 50]$$
$$+ I[Male_i = 1], \text{ and}$$
$$logit(missing_{ij}|Treatment_i = intervention) = -1.8 + I[Male_i = 0].$$

Missingness (under NMAR) was related to time since randomization, treatment arm, and slope for Scenarios A2 and B5:

$$logit(missing_{ij}|Treatment_i = control) = -1.5 + 2 \times I[slope_i \geq 0.2] + time_{ij}, \text{ and}$$
$$logit(missing_{ij}|Treatment_i = intervention) = -2.5 + I[slope_i < 0.5].$$

Missingness (under NMAR) was additionally related to auxiliary terms for Scenarios B6, B7, and C9:

$$logit(missing_{ij}|Treatment_i = control) \quad = -2.5 + 2 \times I[slope_i \geq 0.2] \ + time_{ij}$$
$$+ I[Age_i < 50] + I[Male_i = 1], \text{ and}$$
$$logit(missing_{ij}|Treatment_i = intervention) = -2.5 + I[slope_i < 0.5] + I[Male_i = 0].$$

One thousand simulated data sets were generated for each of nine scenarios.

Modeling approaches considered

A variety of methods were considered (Table 2). For reference, three models were fit: both mixed effects models ((1) the true or RI model for Scenarios A and B, and (2) the full RIRS model or true model for Scenario C that includes both random intercept and random slope) and (3) slope model (also the true model for Scenario C). In addition, the RI, RIRS, and slope models were fit to the incomplete data sets (for the RI and RIRS models, the incomplete data set consisted of any individual who contributed at least one outcome measurement and for the latter model, the incomplete data set included any individual who contributed at least two outcome measurements). The distinction between the slope and RIRS models lies in how the models were fit. In the former, slopes are derived and then linear regression is applied, regressing the derived slope on treatment arm and then using maximum likelihood methods to obtain estimates for the relevant parameter (treatment effect), whereas in the latter, a likelihood that incorporates normally distributed random effects for the slope and intercept is optimized to obtain estimates for the relevant parameter (a time by treatment interaction effect).

Five MI methods – 2 under the active approach, 2 under the passive approach, and 1 hybrid approach – were applied to the slope model. Under *Active imputation (Active)* the slope was derived and imputed as any other variable together with any missing underlying BMI measurements. In addition, slope is only imputed for those missing all 3 follow-up BMI measurements. Inconsistent values can then be obtained if the imputed slope does not match the imputed BMI measurements or when observed slope does not match all BMI measurements in cases where 1 or 2 BMI measurements are imputed. *Active imputation all (Active-*

All) is a variation on *Active* in which slope is only considered observed if all 4 BMI measurements contribute to its derivation. If slope is missing, it is imputed as any other variable as in *Active. Passive* and *Passive-All* involve deriving slope after imputing BMI measurements, where the variable slope can be considered for all variables other than those involved in its derivation. Thus, slope can be used in the imputation of covariates age and gender, for example, but it cannot be used in the imputation of the BMI measurements themselves.

After the imputation step, slope is derived from the underlying BMI measurements. In *Passive* this is done for those missing all 3 follow up measurements, and in *Passive-All* it is done for those missing at least one BMI measurement. As a consequence, internal inconsistencies may still be observed in *Passive* (but not in *Passive-All*) because slope is derived only for a proportion of subjects so that imputed BMI measurements may not match observed slope. The *Derive-Impute-Derive* approach is considered a hybrid approach of active and passive approaches. Slope is derived for those who contribute at least 2 measurements and imputed as any other variable as in *Active* and *Active-All*. After completing the imputation step, however, slope is re-derived from the underlying BMI measurements, some of which are observed and some of which are imputed. By including slope in the imputation model, relevant interrelationships among the variables are preserved. By re-deriving slope after the imputation process, internal consistency is gained. It is unclear, however, whether this step will interfere with what was gained in the imputation step by preserving interrelationships.

Ten imputations ($m=10$) were performed for each MI method for each data set. As passive approaches require imputation be performed using the FCS approach, FCS was used for all MI approaches. Procedures in Stata and R were used for this purpose.

Metrics for evaluation of methods

To evaluate performance of the methods, the relative mean squared error (rMSE) – or the mean squared error (MSE) relative to the MSE for the true model fitted to the full data set (gold standard) – was utilized, because this statistic considers both bias and efficiency. Other relevant statistics to gauge performance included the bias (the difference between the true (-0.5) and estimated parameter averaged over the number of simulations), the average model-based standard error, the empirical standard error, the square root of the variance of the coefficients across the simulations, the coverage probability (the percentage of time the true parameter is contained in the 95% confidence interval constructed using the

model-based standard error), and power (the percentage of time the null hypothesis that the parameter is zero is rejected using a two-sided Wald test at the 0.05 level of significance).

Table 2. Methods considered

Method	Description
RI Model Full	Mixed effects model with random intercept fit to the full data set
Slope Model Full	Slope model fit to the full data set
RIRS Model Full	Mixed effects model with random intercept and slope fit to the full data set
RI Model Incomplete	Mixed effects model with random intercept fit to those who provide at least one outcome measurement
Slope Model Incomplete	Slope model fit to those who provide at least two outcome measurements
RIRS Model Incomplete	Mixed effects model with random intercept and slope fit to those who provide at least one outcome measurement
MI Methods for Slope Model	
Active Imputation	Impute slope as any other variable for those who only have baseline outcome measurement
Active Imputation All	Impute slope as any other variable for those who do not have all four outcome measurements
Passive	Impute BMI using all variables except slope and derive slope for those with only baseline outcome measurements; slope is considered in imputation of variables other than those involved in derivation.
Passive All	Impute BMI using all variables except slope and derive slope for those missing any BMI measurements; slope is considered in imputation of variables other than those involved in derivation
Hybrid Derive-Impute-Derive	Derive slope and use in imputations of all variables and re-derive for consistency after imputation procedure

Results

The gold standard for comparisons is the true model fit to the full data set (RI model for Scenarios A and B and RIRS/slope model for Scenario C). The RIRS model is statistically equivalent to the slope model, although methods for fitting

171

differ – and thus, both the RIRS and slope models are considered the true model for Scenario C.

Performance of methods when data are generated from an RI model and data are MAR (Scenarios A1, B3, and B4)

Performance of RI, Slope, and RIRS models without imputation: When the data were MAR, both the RI model (true model) and the slope model were unbiased (true parameter $= -0.5$) when fit on the observed or incomplete data (e.g., average point estimates for Scenario A1 for all 3 models were -0.498, See Table 3a). The slope model yielded least biased estimates (-0.004 versus -0.005 for the RI and RIRS models, Table 3a). Whereas the RI and RIRS models provided excellent coverage of 95.1% and 94.5 %, however, the slope model provided a lower coverage percentage of 85 (Figure 1). The RI and RIRS models had rMSEs of 6.5 and 6.6, whereas the slope model had an rMSE of 10.2.

Table 3a. Results from applying methods to Scenarios A1-A2

|	Model/Method	Average Beta	Average SE	Bias	Cov	rMSE
*RI, Slope, and RIRS Models Fit to **Full Data from Scenario A** (RI is true model; no auxiliary variables)*	RI (True)	-0.498	0.052	0.002	95.4	1.0
	Slope	-0.498	0.052	0.002	95.4	1.0
	RIRS	-0.498	0.052	0.002	95.9	1.0
A1	*RI, Slope, and RIRS Models Fit to Incomplete Data*					
	RI (True)	-0.505	0.136	-0.005	95.1	6.5
	Slope	-0.504	0.123	-0.004	85.4	10.2
	RIRS	-0.505	0.136	-0.005	94.5	6.6
	Slope Models with MI					
	Active	-0.460	0.130	0.040	87.1	9.9
	Active-All	-0.443	0.583	0.057	99.6	22.2
	Passive	-0.445	0.181	0.055	85.1	15.3
	Passive-All	-0.391	0.275	0.109	77.5	41.5
	Derive-Impute-Derive	-0.447	0.155	0.053	91.4	10.3
A2	*RI, Slope, and RIRS Models Fit to Incomplete Data*					
	RI (True)	-0.260	0.105	0.240	36.5	24.8
	Slope	-0.304	0.114	0.196	56.1	21.9
	RIRS	-0.262	0.106	0.238	38.5	24.4
	Slope Models with MI					
	Active	-0.333	0.120	0.167	67.8	17.3
	Active-All	-0.331	0.443	0.169	94.1	25.2
	Passive	-0.243	0.118	0.257	41.9	28.8
	Passive-All	-0.182	0.144	0.318	37.2	43.4
	Derive-Impute-Derive	-0.331	0.121	0.169	66.3	17.5
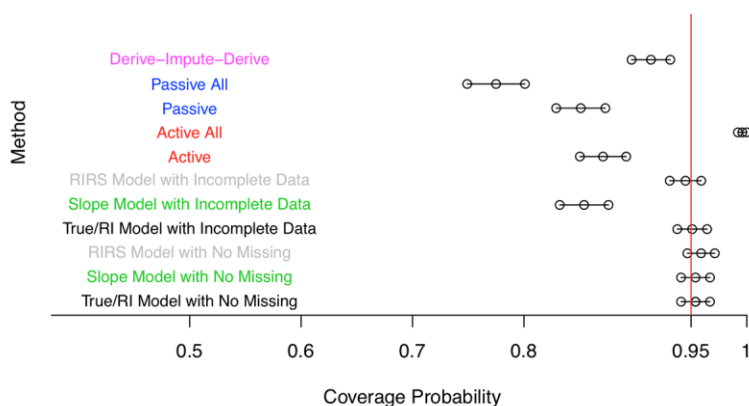
Figure 1. Coverage probability by method for Scenario A1, where data are generated from an RI model, data are MAR, and there are no auxiliary data

Slope model with MI approaches when no auxiliary information is available (Scenario A1): When no auxiliary information was present and the data were MAR (Scenario A1) modest improvement in the slope model could be obtained through use of MI, but depended on the choice of MI method, where the rMSEs ranged from 9.9 (*Active*) to 41.5 (*Passive-All*) and where coverage ranged from 77.5% to 99.6% (Figure 1). Bias ranged from 0.04 (*Active*) to 0.11 (*Passive-All*). The *"All"* approaches suffered a loss in efficiency relative to their corresponding counterparts (e.g., *Active-All* gave an average standard error of 0.58 versus 0.13 for *Active*, and consequently yielded a high coverage of 99.6). With rMSEs of 9.9 and 10.3, the *Active* and *Derive-Impute-Derive* approaches were comparable to the slope model without MI, where *Active* MI provided slight improvement to the slope model without MI. Use of other MI methods with the slope model (*Active-All, Passive, and Passive-All*) gave worse results than using the slope model alone. The true (RI) and RIRS models without imputation performed best among all approaches, when considering bias, average SE, coverage and rMSE.

Slope model with MI approaches when auxiliary information is available (Scenarios B3 and B4): Similar variability across MI methods was observed when auxiliary data were present (rMSEs varied from 8.7 to 20.5 in Scenario B3 and 6.8 to 26.2 in Scenario B4, Table 3b). There was a marked change, however. The inclusion of auxiliary terms in the imputation model greatly affected the performance of passive approaches. To see this, we compared results from

173

Table 3b. Results from applying methods to Scenarios B3-B7

		Model/Method	Avg Beta	Avg SE	Bias	Cov	rMSE
*RI, Slope, and RIRS Models Fit to **Full Data from Scenario B** (RI is true model; no auxiliary variables)*		RI (True)	-0.498	0.052	0.002	94.8	1.0
		Slope	-0.498	0.052	0.002	94.8	1.0
		RIRS	-0.498	0.052	0.002	95.1	1.0
B3	*RI, Slope, and RIRS Models Fit to Incomplete Data*	RI (True)	-0.503	0.136	-0.003	95.5	6.5
		Slope	-0.499	0.123	0.001	84.4	9.9
		RIRS	-0.503	0.136	-0.003	95.6	6.5
	Slope Models with MI	Active	-0.462	0.127	0.038	88.0	9.5
		Active-All	-0.440	0.588	0.060	100.0	20.5
		Passive	-0.456	0.134	0.044	88.8	8.7
		Passive-All	-0.428	0.183	0.072	80.0	20.1
		Derive-Impute-Derive	-0.446	0.154	0.054	90.8	10.3
B4	*RI, Slope, and RIRS Models Fit to Incomplete Data*	RI (True)	-0.530	0.119	-0.030	94.5	5.2
		Slope	-0.481	0.122	0.019	87.9	8.5
		RIRS	-0.531	0.119	-0.031	94.0	5.2
	Slope Models with MI	Active	-0.492	0.127	0.008	90.6	8.1
		Active-All	-0.492	0.698	0.008	99.4	26.2
		Passive	-0.452	0.120	0.048	90.9	6.8
		Passive-All	-0.448	0.158	0.052	88.1	12.3
		Derive-Impute-Derive	-0.488	0.138	0.012	91.1	8.5
B5	*RI, Slope, and RIRS Models Fit to Incomplete Data*	RI (True)	-0.259	0.105	0.241	36.0	24.2
		Slope	-0.310	0.114	0.190	58.2	20.7
		RIRS	-0.262	0.106	0.238	38.2	23.8
	Slope Models with MI	Active	-0.405	0.116	0.095	78.9	11.0
		Active-All	-0.403	0.503	0.097	96.6	20.4
		Passive	-0.368	0.099	0.132	72.1	10.0
		Passive-All	-0.331	0.105	0.169	60.6	14.7
		Derive-Impute-Derive	-0.402	0.117	0.098	79.4	11.3
B6	*RI, Slope, and RIRS Models Fit to Incomplete Data*	RI (True)	-0.297	0.099	0.203	44.9	17.8
		Slope	-0.332	0.115	0.168	64.3	17.0
		RIRS	-0.301	0.100	0.199	47.9	17.2
	Slope Models with MI	Active	-0.426	0.117	0.074	86.0	9.0
		Active-All	-0.431	0.517	0.069	98.3	18.8
		Passive	-0.371	0.098	0.129	72.6	9.4
		Passive-All	-0.339	0.101	0.161	61.0	13.2
		Derive-Impute-Derive	-0.426	0.118	0.074	83.2	9.1
B7	*RI, Slope, and RIRS Models Fit to Incomplete Data*	RI (True)	-0.390	0.064	0.110	59.1	5.8
		Slope	-0.485	0.083	0.015	93.0	2.8
		RIRS	-0.394	0.065	0.106	62.3	5.5
	Slope Models with MI	Active	-0.489	0.083	0.011	93.3	2.8
		Active-All	-0.486	0.498	0.014	100.0	10.2
		Passive	-0.487	0.080	0.013	93.2	2.6
		Passive-All	-0.417	0.063	0.083	73.3	3.9
		Derive-Impute-Derive	-0.490	0.083	0.010	92.6	2.8

Scenario A1 where there were no auxiliary variables to those from Scenarios B3 and B4 (Figure 2). In the latter scenarios, two auxiliary variables incorporated into the imputation process – one continuous and one dichotomous – were related to BMI measurements and the rate of change in BMI, and in Scenario B4 they were additionally related to missingness. *Passive* MI had a performance that was much improved (e.g., rMSE of *Passive* was 8.7 and 6.8 in Scenarios B3 and B4 versus 15.3 in Scenario A1; Figure 2, Table 3b) and provided estimates that were now competitive to those from *Active* MI; *Passive* MI yielded the lowest rMSEs of all MI methods and demonstrated improvement over the slope model without MI (rMSE of 8.7 versus 9.9 and 6.8 versus 8.5 for Scenarios B3 and B4, respectively).

The average standard error by bias across all methods for Scenario B4 is shown in Figure 3. A desirable method is one closest to the (0,0) point on the graph (and closest to the true (RI), RIRS, and slope models fitted to the full data) as this implies low bias and high efficiency. Although *Passive* in the presence of auxiliary terms is more competitive to *Active*, *Active* yields estimates with the least bias (e.g., in Scenario B4, bias for *Active* is 0.008 versus 0.048 for *Passive*). With rMSEs of 8.1 and 6.8, however, neither *Active* nor *Passive* MI applied to the slope model improved overall performance over the true (RI) model without imputation (rMSE=5.2).
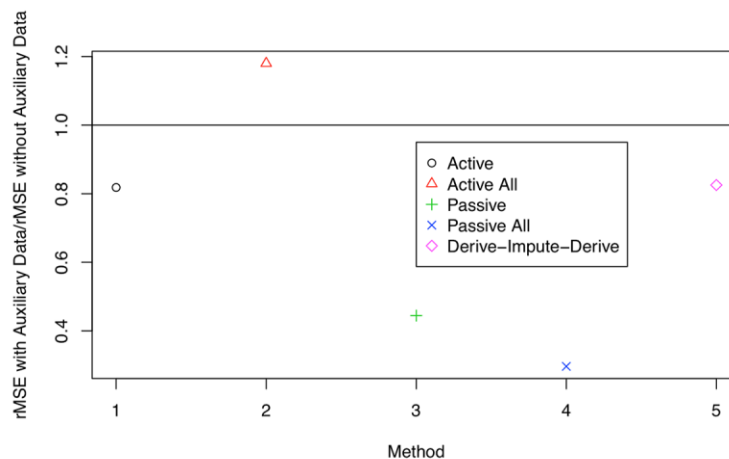

Figure 2. Ratio of relative mean squared error for MI methods with auxiliary data applied to slope model to relative mean squared error for MI methods without auxiliary data applied to slope model
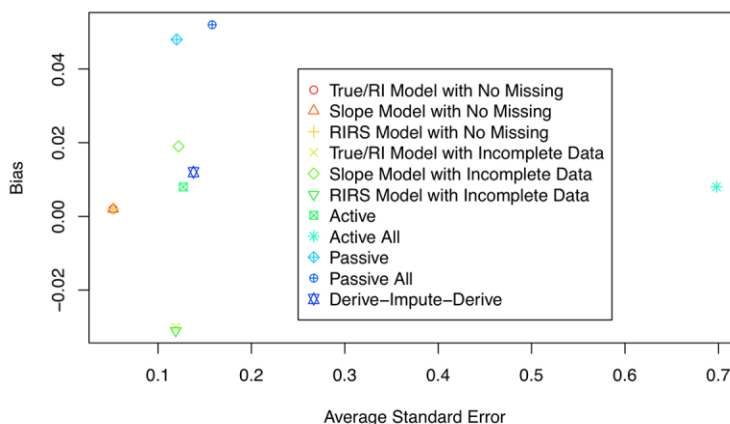
175

Figure 3. Bias by average standard error for Scenario B4, where data, generated from an RI model, are MAR, and auxiliary data are available

Performance of methods when data are generated from an RIRS/slope model and data are MAR (Scenario C8)

Performance of RI, Slope, and RIRS models without imputation: When the data were MAR, among models without imputation, the slope model produced the least biased results (−0.07 for the slope model and −0.097 and −0.092 for the RI and RIRS models; Table 3c). Coverage was lower for the RI model relative to the slope and RIRS models (84.3 versus 89.8 and 90.7 for the slope and RIRS models). The RIRS model gave the lowest rMSE of 3.4 with the RI and slope models providing comparable rMSEs of 3.8 and 4.2.

Slope model with MI approaches when auxiliary information is available (Scenario C8): The rMSEs for the MI approaches applied to the slope model ranged from 3.0 (*Active*) to 8 (*Active-All*). *Active* MI yielded the least biased and most efficient estimates with *Passive* MI performing comparably. The results from fitting a slope model with either *Active* or *Passive* MI indicated improvement over the true models without imputation, where *Active* provided results with the least bias.

176

Table 3c. Results from applying methods to Scenarios C8-C9

		Model/Method	Average Beta	Average SE	Bias	Cov	rMSE
*RI, Slope, and RIRS Models Fit to **Full Data from Scenario C** (RI is true model; no auxiliary variables)*		RI	-0.510	0.069	-0.010	83.8	1.0
		Slope (True)	-0.510	0.096	-0.010	93.9	1.0
		RIRS (True)	-0.510	0.096	-0.010	93.8	1.0
C8	*RI, Slope, and RIRS Models Fit to Incomplete Data*	RI	-0.597	0.138	-0.097	84.3	3.8
		Slope (True)	-0.570	0.157	-0.070	89.8	4.2
		RIRS (True)	-0.592	0.154	-0.092	90.7	3.4
	Slope Models with MI	Active	-0.464	0.143	0.036	89.5	3.0
		Active-All	-0.450	0.693	0.050	99.7	8.0
		Passive	-0.456	0.152	0.044	91.7	3.2
		Passive-All	-0.453	0.196	0.047	90.7	6.0
		Derive-Impute-Derive	-0.461	0.153	0.039	91.4	3.1
C9	*RI, Slope, and RIRS Models Fit to Incomplete Data*	RI	-0.159	0.111	0.341	20.4	14.9
		Slope (True)	-0.253	0.150	0.247	60.7	9.5
		RIRS (True)	-0.258	0.140	0.242	58.7	8.4
	Slope Models with MI	Active	-0.444	0.149	0.056	89.4	3.2
		Active-All	-0.393	0.539	0.107	98.6	6.3
		Passive	-0.403	0.127	0.097	88.9	2.7
		Passive-All	-0.368	0.132	0.132	82.7	3.8
		Derive-Impute-Derive	-0.385	0.134	0.115	83.8	3.6

Performance of methods when data are generated from an RI model and data are NMAR (Scenarios A2, B5, and B6)

Performance of RI, Slope, and RIRS models without imputation: Under NMAR, the slope model yielded results with the least bias, highest coverage and lowest rMSE (Tables 3a and 3b). For example, in Scenario A2, bias ranged from 0.196 (slope model) to 0.24 (RI/true model), coverage ranged from 36.5 (RI/true model) to 56.1 (slope model), and rMSE ranged from 21.9 (slope model) to 24.8 (RI/true model).

Slope model with MI approaches when no auxiliary information is available (Scenario A2:) As under MAR, there was considerable variability among MI methods applied to the slope model with rMSEs ranging from 17.3 to 43.4 (Table 3a). With no auxiliary information, passive approaches exhibited the most bias (0.257 and 0.318 for *Passive* and *Passive-All*) and the worst coverage probabilities (41.9 and 37.2). The active approaches (*Active* and *Active-All*) and

Derive-Impute-Derive proved superior to the passive approaches when considering both bias and efficiency.

Slope model with MI approaches when auxiliary information is available (Scenarios B5 and B6): Improvement in MI methods from incorporating auxiliary data under NMAR was similar to that seen under MAR. Specifically, auxiliary terms greatly improved performances for the passive approaches and modestly for the *Active* and *Derive-Impute-Derive* approaches. All methods, however, had larger improvements when auxiliary variables were also related to missingness (Scenario B6). rMSEs ranged from 10 (*Passive*) to 20.4 (*Active-All*) for Scenario B5 and 9 (*Active*) to 18.8 (*Active-All*) for Scenario B6 and suggested that the performances of *Passive*, *Active*, and *Derive-Impute-Derive* were comparable overall. *Active* and *Derive-Impute-Derive*, however, achieved estimates with comparable and less bias than *Passive* (e.g., 0.095 and 0.098 versus 0.132 in Scenario B5) and higher coverage, where, for example, *Active* and *Derive-Impute-Derive* provided coverage estimates of 86% and 83.2% compared to *Passive*, which yielded coverage of 72.6% in Scenario B6. Figure 4 depicts efficiency and bias across MI methods and shows that under NMAR in the presence of auxiliary information, the true (RI) and slope models performed considerably worse with respect to bias and efficiency than the slope model with (and even without) MI. The *Active* and *Derive-Impute-Derive* approaches were the most preferable choices here (Figure 4).
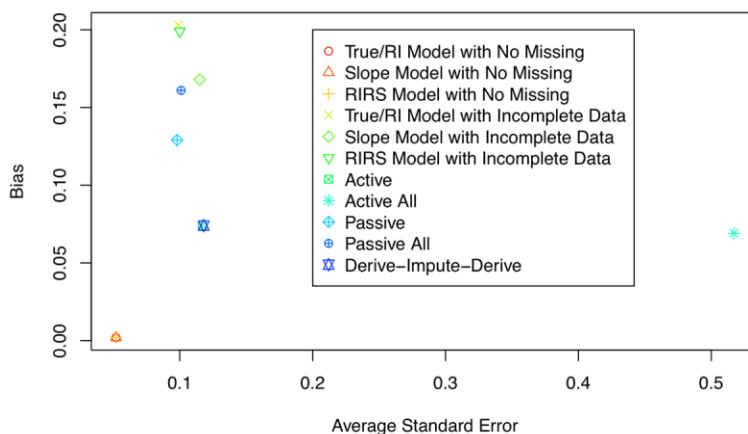

Figure 4. Bias by average standard error for Scenario B6, where data, generated from an RI model, are NMAR, and auxiliary data are available

Performance of methods when data are generated from an RIRS model and data are NMAR (Scenario C9)

Performance of RI, Slope, and RIRS models without imputation: When the data were NMAR, rMSEs ranged from 8.4 (RIRS/true model) to 14.9 (RI model) (Table 3c). Coverage was low for the RI model (20.4%) and higher (although not at nominal levels) for the RIRS and slope models (58.7% and 60%). Bias ranged from 0.242 (RIRS model) to 0.341 (RI model).

Slope model with MI approaches when auxiliary information are available (Scenario C9): When applying MI to the slope model and auxiliary information was present, rMSEs ranged from 2.7 (*Passive*) to 6.3 (*Active-All*) (Table 3c). Although *Passive* yielded the lowest rMSE, *Active* yielded the least biased estimate (0.056 compared to 0.097 for *Passive* and 0.115 for *Derive-Impute-Derive*) and the best coverage of the three competing methods (89.4%). Figure 5 plots the average standard error by the bias for each method. Points corresponding to the *Active*, *Derive-Impute-Derive,* and *Passive* methods were closer to those corresponding to the RI and true (RIRS and slope) models fitted to the full data set than the RI and true (RIRS and slope) models fitted to the incomplete data set, demonstrating the improvement of MI approaches that was not observed under MAR.
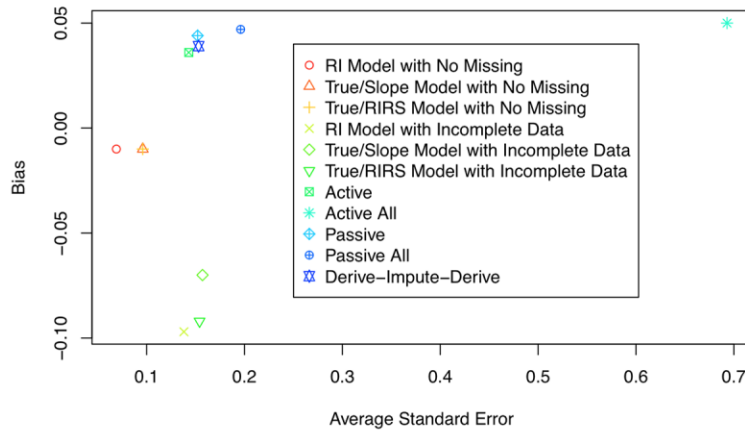

Figure 5. Bias by average standard error for Scenario C8, where data, generated from an RIRS model, are MAR, and auxiliary data are available

Impact of percentage missing under NMAR (comparison of Scenarios B6 and B7)

In Scenario B7, only 20% of the observations were missing on average. Similar patterns emerged in variability across approaches (Table 3b). The slope model without MI was superior to the true model (RI), with relative MSEs of 2.9 versus 5.8, achieved good coverage of 93% compared to only 59.1% for the RI (true) model and 62.3% for the RIRS model. Relative MSEs among MI methods applied to the slope model varied from 2.6 to 10.2. As in Scenario B6 where 40% of the observations were missing, improvement over the slope model without MI was gained with *Active*, *Derive-Impute-Derive*, and *Passive* methods. The advantage, however, was more modest than when the percentage missing was higher.

Conclusion

There was variability in results from applying various MI methods when slope was the outcome and missing for a proportion of subjects, and a two-stage linear regression was applied. When data were generated from an RI model, and the data were MAR, the RI model without imputation was the most appropriate choice with or without auxiliary data. Without auxiliary information, little was gained by doing MI, and in fact, considerable harm could come from using MI under a passive approach. When auxiliary information was present, the performance of *Passive* improved and surpassed that of *Derive-Impute-Derive* and *Active* in terms of the rMSE statistic; all things considered, however, *Active*, *Derive-Impute-Derive*, and *Passive* were comparable. All methods, however, were worse than the true RI model.

When the data were NMAR, much was gained by using a slope model over the RI and RIRS models, and importantly, MI with the slope model was considerably preferable over the slope model alone. When the data arose from either an RI or RIRS model, the slope model without MI outperformed the true model. Even without auxiliary data, *Active* and *Derive-Impute-Derive* with the slope model provided benefit (in terms of MSE and bias) over the models without MI, whereas the passive approaches proved harmful. With auxiliary information, however, *Passive* also provided improved estimates over the models without MI.

Although it is possible that the slope model is more robust to the violation of the MAR condition than the RI model, this difference in performance should be explored under various NMAR conditions, particularly in the absence of auxiliary data. In the presence of auxiliary information, it makes sense that, as we observed,

the slope model with MI outperformed all models fit to the incomplete data without MI because of the incorporation of auxiliary terms, making MAR a more reasonable assumption, and therefore a better choice under such conditions.

There were two *All* approaches considered, which were variations on the *Active* and *Passive* methods: *Active-All*, and *Passive-All*. Each performed worse than its counterpart. The idea behind the *All* approaches was that better slope measurements could be obtained by providing one based on a complete set of values (in this case, 4) rather than one based on only a partially observed trajectory, which may then yield less biased estimates of the relationship of interest.

A reduction in bias was not observed, however. Instead, the *All* approaches were more inefficient and more biased than their counterparts, where *Active-All* in particular suffered the largest loss in efficiency. The loss in efficiency may be explained by the application of these approaches to a modified data set with a higher proportion of missing data. More specifically, the *All* approaches assumed slope was missing for those missing at least 1 BMI measurement as opposed to for those missing all 3 follow-up BMI measurements, thereby increasing the fraction of missing information. Despite the increase in missing observations, the *All* method was applied with the same number of imputations ($m=10$) as with its counterpart method. Even when the number of imputations was doubled, however, these relative losses in efficiency remained (data not shown).

It may be that a more substantial increase in the number of imputations is needed, and this remains to be explored. In the active setting, such an objective simply adds a new variable to be imputed, while making use of the partially observed slope as an auxiliary term. While the objective of the passive approaches is to have internal consistency, full consistency was not achieved in *Passive*, because the observed slope and not the derived slope was used for subjects with at least 2 BMI outcome measurements. Although full internal consistency was achieved in *Passive-All* gains in properties of the estimates were not observed. These approaches likely failed because the focus was incorrectly on producing imputed values closer to what would have been observed (i.e., by making them consistent) with the idea that such values would lead to better descriptions of relationships rather than on exploiting the interrelationships among the variables. It was demonstrated that use of observed slope (even if based on only a fraction of measurements) produced estimates of relevant parameters with superior properties to those based on a full number of measurements where a fraction of those measurements were imputed. Findings may vary, however, in settings with a larger number of longitudinal measurements.

Derive-Impute-Derive behaved comparably to *Active* MI in all scenarios including in the presence or absence of auxiliary information. An active-type approach that requires an extra step of editing or re-deriving for consistency, *Derive-Impute-Derive* did not demonstrate meaningful improvements over *Active*. Larger variation in performance may be observed in settings with higher proportions of internal consistency. This would need to be explored in further simulation studies.

Variability between *Active*/*Derive-Impute-Derive* MI and passive approaches was largely due to the presence of auxiliary information. Although some improvement was observed with *Active* and *Derive-Impute-Derive* with the presence of auxiliary information, it was modest relative to the improvement observed among the passive approaches. It was the auxiliary information that allowed *Passive* to be competitive to *Active* and *Derive-Impute-Derive;* because observed slope is not considered an auxiliary term when passively imputing BMI, other auxiliary information is critical for gleaning the interrelationships among relevant variables in the model. In contrast, in *Active* and *Derive-Impute-Derive*, BMI measurements (and other variables) are used as auxiliary terms when imputing slope and vice versa and the interrelationships among these terms are jointly considered. It would be interesting to further explore the differences in performance of approaches in settings where other covariates have missing values.

There were some computational issues to consider when implementing the passive and hybrid approaches. Unlike *Active*, they required some extra steps beyond the usual MI programming. *Derive-Impute-Derive* requires re-deriving slope to ensure internal consistency after doing an active-type imputation (See Appendix A for STATA code). *Passive* – possible in STATA and R, but not in SAS – requires an explicit statement to define the derived term as a function of the main effects so that updating of the derived variable can be performed during the imputation procedure (See Appendix B). With a derived term that requires more than a basic operation of addition or multiplication as in interaction effects, this can be challenging. If the derivation requires more than one line, this may be problematic in STATA. Thus, if performances across MI methods are equal, *Active* is preferred for its computational simplicity. Further, it is important to note that *Active* – unlike *Passive* – is available in all MI packages including R, STATA, SAS and SPSS.

It may be argued that the imputation methods considered here do not lead to proper imputations. Rubin provides criteria required (i.e., properties necessary for the imputation method) to yield imputations that are deemed proper (Rubin, 1987). Consequently, estimates of population quantities that incorporate proper multiple

imputations are statistically valid (unbiased with appropriate coverage). One such assumption requires congeniality of the imputation and scientific models, where congeniality implies that the imputed data can be derived from the scientific model (Meng, 1994).

An example of an un-congenial model may be one where the scientific model is a linear regression of BMI on treatment arm and age, where BMI is missing for a proportion of subjects, and the imputation model assumes multivariate normality of BMI, treatment arm, and possibly other auxiliary terms, but excludes age. Because the scientific model considers interrelationships between age and BMI and treatment arm, and the imputation model does not, the models are said to be un-congenial. In this case, when the slope model is the scientific model, recall that there are two stages involved: the derivation of the slopes (Equation 1a) followed by the regression of the derived slopes on treatment arm (Equation 1b).

Because the imputation model does not incorporate uncertainty of the parameters in Equation 1a, it may be considered un-congenial. Imputing under a fully Bayesian paradigm that puts priors on all the parameters listed in both Equations 1a and 1b when deriving the posterior predictive distribution would be considered proper. MI methods considered here, however, were only those that could be implemented using mainstream software like Stata, R, and SAS, with minimal additional code, and specifically where the slope model was the scientific model of interest. Our reference for ideal performance was the true model fit to the data set with no missing values, rather than the ideal MI method.

Further, methods that incorporate MI with mixed effects models such as PAN, developed by Schafer (Schafer, 1997a; Schafer, 1997b), were not considered. Nor was a shared parameter model considered, which is also appropriate for longitudinal models and allows the missing data mechanism to be NMAR (e.g., Wu & Bailey, 1989). The former approach can be implemented using a package in R, whereas code for fitting the shared parameter model is not available in mainstream software packages. It would be interesting to compare MI methods under both the slope model and mixed effects models in the future. The current focus, however, was specifically on fitting the slope model, where the mixed effects models fit to the full and incomplete data without MI served as references. Clinical investigators are often averse to applying mixed effects models if they can apply a seemingly simpler two-stage linear regression model. This is the case with Stanford GOALS. Thus, the intent was to evaluate practical MI methods under the slope model.

The goal was to provide a set of guidelines on how to analyze rate of change in the presence of missing data. Figure 6 provides a flow chart to aid in decision-making. Typically, the first step when faced with missing data is to consider plausible missing data mechanisms.

We take one more step back, however, and ask to consider from which model the data may be reasonably generated. In this study, data were generated from both an RI model and the RIRS model. The former assumes each individual has a BMI around which a series of BMI values center. Further it assumes that how those BMI values change over time may be affected by treatment regimen. Data arising from an RIRS model implies each individual additionally has his/her own rate of change in BMI, and, as before, trajectories may be affected by a treatment regimen. It leads to slight differences in recommendations. If the data are likely to arise from a slope model (i.e., if you can assume BMI values arise from a rate of change), we recommend using a slope model with *Active* MI (if either MAR or NMAR, and in the presence of auxiliary information).

Although some improvements were observed with *Passive* over *Active* when auxiliary data were present, *Active* was still less biased, comparably efficient and simpler to implement. If auxiliary information are not available, a more sophisticated approach should be considered such as the shared parameter model as nominal coverage was not achieved with the methods considered here (Duch & Stevenson, 2005). If the data arise from an RI model, and if the data are MAR, the best choice among those considered is an RI model. Thus, investigators should reconsider specifying the scientific model as the slope model in favor of the RI model; the slope model without MI produced estimates with inferior properties. Further, in the absence of auxiliary information, little was gained with using MI. Some investigators, however, may be wary to fit and interpret parameters from a mixed effects model and may wish for the simplicity of a linear regression model (e.g., Liao et al., 2013) the benefits of which are described here (Rogosa & Saner, 1995; Gelman, 2005). In this case, *Active* MI can improve upon the slope model without MI in the presence of auxiliary information. If the data are NMAR, the slope model with *Active* MI provides improvement over the mixed effects models in the presence of auxiliary terms. In general, however, if the data are NMAR, a more sophisticated approach such as the shared parameter model should be considered.
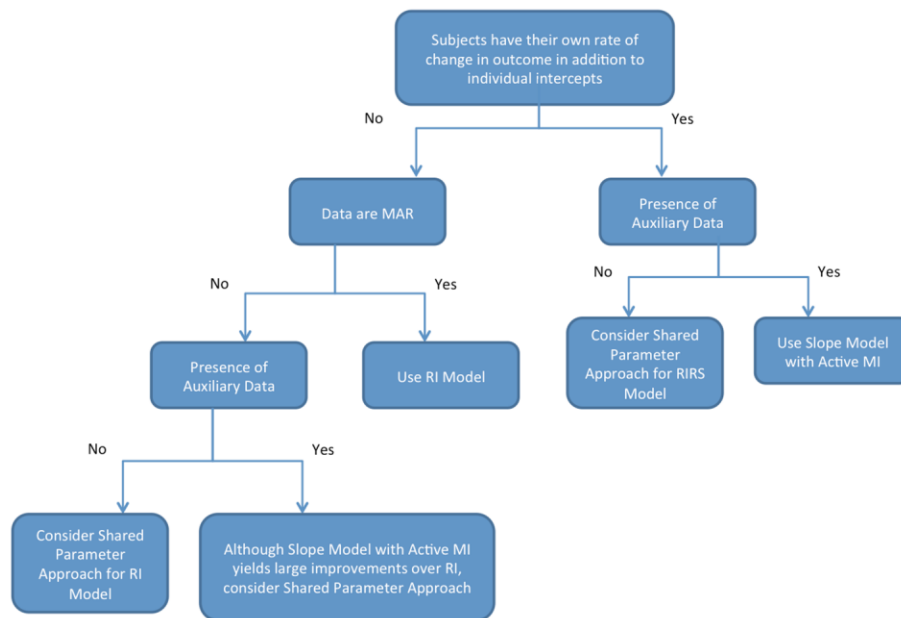
Figure 6. Flow chart of model choices

In longitudinal clinical trial settings such as Stanford GOALS it is likely the case that the data are NMAR. For example, those who respond well to the behavioral intervention may be more likely to have outcome values and intermittent BMI values. The availability of numerous solid auxiliary variables, however, is also likely, and conditional on these, it may be reasonable to assume the data are MAR. Baseline measurements of BMI, other adiposity measures, lipid profiles, as well as intermittent values of BMI are potentially useful auxiliary terms that can be incorporated into the analysis. Thus, if NMAR, standard MI methods that incorporate such terms can provide benefit over the mixed effects and slope models without MI. Although it is important to pre-specify a primary method of analyses (in Stanford GOALS, the primary approach is the slope model with *Active* MI), it is recommended that sensitivity analyses be performed – in particular, the mixed effects and slope models without MI, and the slope model with *Derive-Impute-Derive*, and *Passive* approaches. If results are comparable, this bolsters the findings from the primary analysis. If results are heterogeneous, it is critical to report the discrepancy in order to provide a context for the audience in which to interpret the findings.

Acknowledgements

References

Allison, P. D. (2002). *Missing data*. Thousand Oaks, CA: Sage.

Collins, L. M., Schafer, J. L., & Kam, C.M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, *6*(4), 330-51. doi: 10.1037/1082-989X.6.4.330

Duch, R. M. and Stevenson, R. (2005). Context and the economic vote: a multilevel analysis. *Political Analysis*, *13*(4), 387-409. doi: 10.1093/pan/mpi028

Gelman, A. (2005). Two-stage regression and multilevel modeling: a commentary. *Political Analysis*, *13*(4), 459-461. doi: 10.1093/pan/mpi032

Jusko, K. L. and Shively, W. P. (2013). Applying a two-step strategy to the analysis of cross-national public opinion data. *Political Analysis*, *13*(4), 327-344. doi: 10.1093/pan/mpi030

Liao, J., Brunner, E. J., & Kumari, M. (2013). Is there an association between work stress and diurnal cortisol patterns? Findings from the Whitehall II study. *PLoS One*, *8*(12). doi: 10.1371/journal.pone.0081020

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. 2nd ed. New York: J. Wiley & Sons.

Meng, X-L. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statistical Science, 9*, 538-558.

Mitani, A., Kurian, A., Das, A., & Desai, M. (2015). Navigating choices when applying multiple imputation in the presence of multi-level categorical interaction effects. *Statistical Methodology*, *27*, 82-99. doi: 10.1016/j.stamet.2015.06.001

Robinson, T. N., Matheson, D., Desai, M., Wilson, D. M., Weintraub, D. L., Haskell, W. L., McClain, A., McClure, S., Banda, J. A., Sanders, L. M., Haydel, K. F., & Killen, J. D. (2013). Family, community and clinic collaboration to treat overweight and obese children: Stanford GOALS − A randomized controlled trial of a three-year, multi-component, multi-level, multi-setting intervention. *Contemporary Clinical Trials, 36*(2), 421-435. doi: 10.1016/j.cct.2013.09.001

Rogosa, D. & Saner, H. (1995). Longitudinal data analysis examples with random coefficient models. *Journal of Educational and Behavioral Statistics*, *20*(2), 149-170. doi: 10.3102/10769986020002149

Royston, P. (2009). Multiple imputation of missing values: further update of ice, with an emphasis on categorical variables. *Stata Journal*, 9(3), 466-77.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, *91*(434), 473-89. doi: 10.2307/2291635

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: J. Wiley & Sons.

Schafer, J. L. (1997a). *Imputation of missing covariates under a general linear mixed model* (Technical Report). State College, PA: Dept. of Statistics, Penn State University.

Schafer, J. L. (1997b). *Analysis of incomplete multivariate data*. London: Chapman & Hall.

Seaman, S. R., Bartlett, J. W., White, I. R. (2012). Multiple imputation of missing covariates with non-linear effects and interactions: an evaluation of statistical methods. *BMC Medical Research Methodology, 12*:46. doi: 10.1186/1471-2288-12-46

von Hippel, P. T. (2009). How to impute interactions, squares and other transformed variables. *Sociological Methodology*, *39*, 265-91.

van Buuren, S. (2012). *Flexible imputation of missing data*. London: Chapman & Hall/CRC.

van Buuren S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research*, *16*(3), 219-42. doi: 10.1177/0962280206074463

White, I. R., Royston, P., Wood, A. M. (2011). Multiple imputation using chained equations: issues and guidance for practice. *Statistics in Medicine, 30*(4), 377-99. doi: 10.1002/sim.4067

Wu, M. C. & Bailey, K. R. (1989). Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics*, *45*, 939-955.

Appendix A: *Derive-Impute-Derive* Imputation of Slope Using ICE

```
* Read in ith data set
insheet using "/Users/mdesai/Documents/COPTR/Multiple Imputation
Research/Simulated Data Sets/reducedwidescenA`i'.csv", clear

destring, ignore("NA") replace

* Derive and Impute:
* Doing Active imputation by including slope in the imputation model as any
variable

ice reducedslope bmi0 bmi1 bmi2 bmi3 txgroup age male,
saving("/Users/mdesai/Documents/nimpute.dta") m(10) replace

use "/Users/mdesai/Documents/nimpute.dta", clear

* Reshape data set from wide format to long format so that slope can be re-
derived
reshape long bmi, i(id _mj) j(visit)

* Derive again
statsby _b, by(id _mj) saving("/Users/mdesai/Documents/newb.dta", replace): reg
bmi visit

* Reshape from long to wide format and sort for analysis
reshape wide bmi, i(id _mj) j(visit)

use "/Users/mdesai/Documents/newb.dta", clear
sort id _mj

use "/Users/mdesai/Documents/nimpute.dta", clear
sort id _mj

merge 1:1 id _mj using "/Users/mdesai/Documents/newb.dta"

save "/Users/mdesai/Documents/merge.dta", replace

use "/Users/mdesai/Documents/merge.dta", clear

sort _mj id

* Now fit the scientific model using the re-derived outcome and summarize using
MICOMBINE
* where the newly derived term is called _b_visit

micombine reg _b_visit txgroup
```

189

Appendix B: *Passive* and *Passive-All* Imputation of Slope Using ICE

```
* read in the ith data set

insheet using "/Users/mdesai/Documents/COPTR/Multiple Imputation
Research/Simulated Data Sets/reducedwidescenA`i'.csv", clear

destring, ignore("NA") replace

* copy reducedslope
gen newslope = reducedslope

* passively impute reducedslope using ICE

ice bmi0 bmi1 bmi2 bmi3 txgroup age male newslope, passive(newslope: (-
1.5*(bmi0-(bmi0+bmi1+bmi2+bmi3)/4) + -0.5*(bmi1-(bmi0+bmi1+bmi2+bmi3)/4) +
0.5*(bmi2-(bmi0+bmi1+bmi2+bmi3)/4) + 1.5*(bmi3-(bmi0+bmi1+bmi2+bmi3)/4))/5)
saving(/Users/mdesai/Documents/nimpute.dta) m(10) seed(1234) replace
use "/Users/mdesai/Documents/nimpute.dta", clear

* Fit scientific model for Passive-All and summarize findings using MICOMBINE.
* Since newslope is used, it replaces the slope based on partially observed BMI
values

micombine reg newslope txgroup

*Fit scientific model for Passive and summarize findings using MICOMBINE.
*Here the derived slope is only used when all 3 BMI measurements are missing.

use "/Users/mdesai/Documents/nimpute.dta",clear
replace reducedslope = newslope if missing(reducedslope)
micombine reg reducedslope txgroup
```

Appendix C: Full Results from Applying Methods to Scenarios A1-C9 (Table A1)

		Model/ Method	Avg Beta	Avg SE	Emp SE	Avg Bias	Cov	MSE	rMSE	Power
	RI, Slope, and RIRS Models Fit to ***Full Data from Scenario A*** *(RI is true model; no auxiliary variables)*	RI (True)	-0.498	0.052	0.053	0.002	95.4	2.8	1.0	100.0
		Slope	-0.498	0.052	0.053	0.002	95.4	2.8	1.0	100.0
		RIRS	-0.498	0.052	0.053	0.002	95.9	2.8	1.0	100.0
A1	*RI, Slope, and RIRS Models Fit to Incomplete Data*	RI (True)	-0.505	0.136	0.134	-0.005	95.1	18.0	6.5	95.9
		Slope	-0.504	0.123	0.167	-0.004	85.4	28.0	10.2	93.9
		RIRS	-0.505	0.136	0.135	-0.005	94.5	18.1	6.6	96.1
	Slope Models with MI	Active	-0.460	0.130	0.160	0.040	87.1	27.3	9.9	89.0
		Active-All	-0.443	0.583	0.241	0.057	99.6	61.3	22.2	1.3
		Passive	-0.445	0.181	0.198	0.055	85.1	42.2	15.3	72.2
		Passive-All	-0.391	0.275	0.320	0.109	77.5	114.4	41.5	34.4
		Derive-Impute-Derive	-0.447	0.155	0.160	0.053	91.4	28.3	10.3	80.7
A2	*RI, Slope, and RIRS Models Fit to Incomplete Data*	RI (True)	-0.260	0.105	0.104	0.240	36.5	68.5	24.8	69.2
		Slope	-0.304	0.114	0.148	0.196	56.1	60.3	21.9	69.3
		RIRS	-0.262	0.106	0.104	0.238	38.5	67.2	24.4	69.9
	Slope Models with MI	Active	-0.333	0.120	0.141	0.167	67.8	47.7	17.3	74.1
		Active-All	-0.331	0.443	0.202	0.169	94.1	69.6	25.2	0.5
		Passive	-0.243	0.118	0.115	0.257	41.9	79.5	28.8	54.7
		Passive-All	-0.182	0.144	0.136	0.318	37.2	119.6	43.4	26.1
		Derive-Impute-Derive	-0.331	0.121	0.140	0.169	66.3	48.2	17.5	74.3
	RI, Slope, and RIRS Models Fit to ***Full Data from Scenario B*** *(RI is true model; no auxiliary variables)*	RI (True)	-0.498	0.052	0.053	0.002	94.8	2.8	1.0	100.0
		Slope	-0.498	0.052	0.053	0.002	94.8	2.8	1.0	100.0
		RIRS	-0.498	0.052	0.053	0.002	95.1	2.8	1.0	100.0
B3	*RI, Slope, and RIRS Models Fit to Incomplete Data*	RI (True)	-0.503	0.136	0.136	-0.003	95.5	18.4	6.5	95.1
		Slope	-0.499	0.123	0.168	0.001	84.4	28.2	9.9	92.9
		RIRS	-0.503	0.136	0.136	-0.003	95.6	18.4	6.5	95.2
	Slope Models with MI	Active	-0.462	0.127	0.157	0.038	88.0	26.1	9.5	91.3
		Active-All	-0.440	0.588	0.230	0.060	100.0	56.5	20.5	0.7
		Passive	-0.456	0.134	0.150	0.044	88.8	24.4	8.7	90.1
		Passive-All	-0.428	0.183	0.226	0.072	80.0	56.4	20.1	66.4
		Derive-Impute-Derive	-0.446	0.154	0.159	0.054	90.8	28.3	10.3	81.3
B4	*RI, Slope, and RIRS Models Fit to Incomplete Data*	RI (True)	-0.530	0.119	0.118	-0.030	94.5	14.7	5.2	99.2
		Slope	-0.481	0.122	0.155	0.019	87.9	24.2	8.5	94.3
		RIRS	-0.531	0.119	0.118	-0.031	94.0	14.8	5.2	99.2
	Slope Models with MI	Active	-0.492	0.127	0.151	0.008	90.6	22.7	8.1	95.0
		Active-All	-0.492	0.698	0.271	0.008	99.4	73.3	26.2	0.9
		Passive	-0.452	0.120	0.130	0.048	90.9	19.1	6.8	94.4
		Passive-All	-0.448	0.158	0.178	0.052	88.1	34.5	12.3	79.0
		Derive-Impute-Derive	-0.488	0.138	0.154	0.012	91.1	23.9	8.5	92.9

Table A1 continued

		Model/ Method	Avg Beta	Avg SE	Emp SE	Avg Bias	Cov	MSE	rMSE	Power
B5	*RI, Slope, and RIRS Models Fit to Incomplete Data*	RI (True)	-0.259	0.105	0.105	0.241	36.0	68.9	24.2	70.0
		Slope	-0.310	0.114	0.152	0.190	58.2	58.9	20.7	72.5
		RIRS	-0.262	0.106	0.106	0.238	38.2	67.6	23.8	70.2
	Slope Models with MI	Active	-0.405	0.116	0.147	0.095	78.9	30.4	11.0	88.5
		Active-All	-0.403	0.503	0.218	0.097	96.6	57.0	20.4	0.4
		Passive	-0.368	0.099	0.103	0.132	72.1	28.0	10.0	94.7
		Passive-All	-0.331	0.105	0.112	0.169	60.6	41.2	14.7	87.1
		Derive-Impute-Derive	-0.402	0.117	0.148	0.098	79.4	31.6	11.3	86.5
B6	*RI, Slope, and RIRS Models Fit to Incomplete Data*	RI (True)	-0.297	0.099	0.096	0.203	44.9	50.6	17.8	86.5
		Slope	-0.332	0.115	0.143	0.168	64.3	48.4	17.0	76.1
		RIRS	-0.301	0.100	0.098	0.199	47.9	48.9	17.2	87.1
	Slope Models with MI	Active	-0.426	0.117	0.140	0.074	86.0	25.2	9.0	92.7
		Active-All	-0.431	0.517	0.219	0.069	98.3	52.7	18.8	1.2
		Passive	-0.371	0.098	0.099	0.129	72.6	26.4	9.4	96.9
		Passive-All	-0.339	0.101	0.105	0.161	61.0	37.0	13.2	89.8
		Derive-Impute-Derive	-0.426	0.118	0.142	0.074	83.2	25.6	9.1	91.5
B7	*RI, Slope, and RIRS Models Fit to Incomplete Data*	RI (True)	-0.390	0.064	0.066	0.110	59.1	16.4	5.8	100.0
		Slope	-0.485	0.083	0.089	0.015	93.0	8.1	2.8	99.9
		RIRS	-0.394	0.065	0.066	0.106	62.3	15.6	5.5	100.0
	Slope Models with MI	Active	-0.489	0.083	0.088	0.011	93.3	7.9	2.8	99.9
		Active-All	-0.486	0.498	0.168	0.014	100.0	28.4	10.2	2.4
		Passive	-0.487	0.080	0.085	0.013	93.2	7.4	2.6	99.9
		Passive-All	-0.417	0.063	0.064	0.083	73.3	11.0	3.9	100.0
		Derive-Impute-Derive	-0.490	0.083	0.089	0.010	92.6	7.9	2.8	99.9
	*RI, Slope, and RIRS Models Fit to **Full Data from Scenario C** (RI is true model; no auxiliary variables)*	RI	-0.510	0.069	0.096	-0.010	83.8	9.3	1.0	100.0
		Slope (True)	-0.510	0.096	0.096	-0.010	93.9	9.3	1.0	100.0
		RIRS (True)	-0.510	0.096	0.096	-0.010	93.8	9.3	1.0	100.0
C8	*RI, Slope, and RIRS Models Fit to Incomplete Data*	RI	-0.597	0.138	0.161	-0.097	84.3	35.3	3.8	97.4
		Slope (True)	-0.570	0.157	0.185	-0.070	89.8	39.2	4.2	90.7
		RIRS (True)	-0.592	0.154	0.153	-0.092	90.7	31.8	3.4	96.9
	Slope Models with MI	Active	-0.464	0.143	0.164	0.036	89.5	28.1	3.0	85.9
		Active-All	-0.450	0.693	0.268	0.050	99.7	74.5	8.0	0.3
		Passive	-0.456	0.152	0.166	0.044	91.7	29.5	3.2	83.8
		Passive-All	-0.453	0.196	0.231	0.047	90.7	55.4	6.0	65.8
		Derive-Impute-Derive	-0.461	0.153	0.166	0.039	91.4	29.1	3.1	82.2
C9	*RI, Slope, and RIRS Models Fit to Incomplete Data*	RI	-0.159	0.111	0.147	0.341	20.4	137.9	14.9	36.5
		Slope (True)	-0.253	0.150	0.164	0.247	60.7	87.9	9.5	40.6
		RIRS (True)	-0.258	0.140	0.138	0.242	58.7	77.5	8.4	46.8
	Slope Models with MI	Active	-0.444	0.149	0.164	0.056	89.4	30.0	3.2	81.2
		Active-All	-0.393	0.539	0.217	0.107	98.6	58.4	6.3	0.3
		Passive	-0.403	0.127	0.127	0.097	88.9	25.5	2.7	89.4
		Passive-All	-0.368	0.132	0.133	0.132	82.7	35.0	3.8	79.0
		Derive-Impute-Derive	-0.385	0.134	0.143	0.115	83.8	33.7	3.6	80.6