Statistical Methodology 27 (2015) 82–99

Contents lists available at ScienceDirect

Statistical Methodology

journal homepage: www.elsevier.com/locate/stamet

Navigating choices when applying multiple imputation in the presence of multi-level categorical interaction effects

Aya A. Mitani [a,*], Allison W. Kurian [b,1], Amar K. Das [c], Manisha Desai [d,2]

[a] *Department of Biostatistics, Boston University, 801 Massachusetts Avenue, Boston, MA 02118, United States*

[b] *Department of Medicine and Department of Health Research and Policy, Stanford University School of Medicine, 900 Blake Wilbur, Stanford, CA 94305, United States*

[c] *Department of Psychiatry and The Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine, 1 Rope Ferry Rd, Hanover, NH 03755, United States*

[d] *Quantitative Sciences Unit, Department of Medicine, Stanford University School of Medicine, 1070 Arastradero Road, Palo Alto, CA 94304, United States*

ARTICLE INFO

Article history:
Received 18 March 2014
Received in revised form
13 May 2015
Accepted 1 June 2015
Available online 22 June 2015

Keywords:
Multiple imputation
Categorical variables
Interaction effects
Passive imputation
Active imputation

ABSTRACT

Multiple imputation (MI) is an appealing option for handling missing data. When implementing MI, however, users need to make important decisions to obtain estimates with good statistical properties. One such decision involves the choice of imputation model – the joint modeling (JM) versus fully conditional specification (FCS) approach. Another involves the choice of method to handle interactions. These include imputing the interaction term as any other variable (active imputation), or imputing the main effects and then deriving the interaction (passive imputation). Our study investigates the best approach to perform MI in the presence of interaction effects involving two categorical variables. Such effects warrant special attention as they involve multiple correlated parameters that are handled differently under JM and FCS modeling. Through an extensive simulation study, we compared active, passive and an improved passive approach under FCS, as JM precludes passive imputation. We additionally compared JM and

* Corresponding author.
E-mail addresses: amitani@bu.edu (A.A. Mitani), akurian@stanford.edu (A.W. Kurian), amar.das@dartmouth.edu (A.K. Das), manisha.desai@stanford.edu (M. Desai).

1 Tel.: +1 650 498 6004.

2 Tel.: +1 650 725 1946.

http://dx.doi.org/10.1016/j.stamet.2015.06.001

FCS techniques using active imputation. Performance between active and passive imputation was comparable. The improved passive approach proved superior to the other two particularly when the number of parameters corresponding to the interaction was large. JM without rounding and FCS using active imputation were also mostly comparable, with JM outperforming FCS when the number of parameters was large. In a direct comparison of JM active and FCS improved passive, the latter was the clear winner. We recommend improved passive imputation under FCS along with sensitivity analyses to handle multi-level interaction terms.

1. Introduction

Multiple imputation (MI) is an increasingly popular approach for handling missing data [7,8,11,19]. Largely this is due to a growing awareness of the potential bias and inefficiencies that result from applying inappropriate methods, and an increase in software accessibility to perform MI [28]. For example, mainstream software packages such as SAS/STAT software [20], Stata [22], and R [16] offer MI-based analyses. Despite this, a complete-case (CC) analysis, which restricts the analysis to observations with no missing values, still remains the most commonly applied approach, perhaps because it is the default option for handling missing data in all statistical software packages [10,23]. CC analysis, however, is valid when the data are missing completely at random (MCAR) (i.e., missingness is not related to observed or unobserved features), an assumption that does not typically hold in practice. If violated, CC analysis can result in biased and inefficient estimates. MI, on the other hand, is statistically valid under a more flexible assumption about the missing data mechanism; it relies on the assumption that the data are missing at random (MAR) or that missingness is related to observed features only (i.e., after conditioning on relevant observed features, missingness is unrelated to unobserved values). Briefly, MI is a simulation-based approach for filling in each missing datum with a plausible value repeatedly to account for the uncertainty of the sampled values and the imputation process itself. It requires the specification of two statistical models: an imputation model, which is used to impute the missing data for m imputed datasets, and a scientific model, which is used to analyze each of the m imputed datasets in order to address the research question [18].

In addition to being the default approach in software packages, CC analysis may be preferred for its simplicity. Another possible barrier to incorporating MI in the analysis is the numerous choices faced by analysts when implementing MI. Importantly, these choices can have great impact on the results. Among the various choices are the specification of the imputation model (i.e., which variables to consider in the imputation model and their functional form) [5], and the imputation approach. The two main imputation approaches are the joint modeling (JM) approach and the fully conditional specification (FCS) approach. Briefly, JM involves specifying a joint distribution for the data, which is typically assumed to be multivariate normal, in order to derive the posterior predictive distribution from which to impute values [24]. FCS bypasses the specification of a joint model and instead directly specifies the conditional distribution for each partially observed variable [24]. The latter may present advantages for data that contain variables of mixed type, such as binary and categorical variables, where specifying a joint distribution for the data is particularly challenging. While the theoretical properties of estimates generated by JM are well established [14], they are less tractable for FCS, although its use has been well justified empirically through simulation studies [26,29]. In his comprehensive review of these two methods, van Buuren compares and contrasts performance of these two methods. He recommends JM when a multivariate normal assumption is sensible and FCS in the presence of variables of mixed type [24].

Another choice posed to analysts involves how to impute derived variables such as interaction terms. There are two main approaches for handling interaction terms. One approach is to transform

Table 1
Differences between predictors of active imputation and passive imputation approaches in the imputation model under FCS.

Scientific model: $Y = X_1 + X_2 + X_1X_2$			
Variable with missing values	Predictors in imputation model		
	Active	Passive	Improved passive
X_1	Y, X_2, X_1X_2, Z	Y, X_2, Z	Y, X_2, YX_2, Z
X_2	Y, X_1, X_1X_2, Z	Y, X_1, Z	Y, X_1, YX_1, Z
$X_1X_2: X_1 \times X_2$	Y, X_1, X_2, Z	–	–
Z	Y, X_1, X_2, X_1X_2	Y, X_1, X_2, X_1X_2	Y, X_1, X_2, X_1X_2

Z: Auxiliary variable.
YX_1: Interaction between Y and X_1.
YX_2: Interaction between Y and X_2.

or derive the variable (by taking the product of the corresponding main effects) and then to impute it as with any other variable (*transform–impute*). Another approach is to impute the data and then to subsequently transform or derive the interaction variable (*impute–transform*) [27]. The former is referred to as *active imputation*, and can be applied under either JM or FCS approach. However, this approach is likely to produce an imputed interaction term that is not the product of its corresponding main effects. The latter, on the other hand, has a seemingly desirable property of preserving a consistent relationship between the imputed interaction term and its corresponding main effects. However, the imputation model may not be compatible with the scientific model that describes the outcome as a function of interactions [15,19]. A variation of this approach that partially addresses this problem is to include the interaction variable in the imputation model, but later replace the imputed interaction term with the product of the imputed main effects (*transform–impute–transform*) in order to preserve the correct relationship. This method also can be implemented under either JM or FCS approach. An iterative variation of this approach that can be implemented under FCS only is called *passive imputation*. Specifically, to passively impute an interaction term, one first computes the interaction term and considers it in the imputation of all variables with the exception of those used to derive the interaction (Table 1). Missing values of the interaction term are then derived (not imputed) as a function of the main effects [17]. This is more desirable than the *impute–transform* method, as the interaction term is still considered in the imputation process with the benefit of achieving internal consistency between the interaction effect and its corresponding main effects. Another method referred to as *improved passive imputation* has recently been introduced, where the interaction of one of the main effects and the outcome is included as a predictor when imputing the other main effect (Table 1). Intuition behind why this approach may reduce bias compared to the conventional *passive imputation* is that if there were a true interaction between two main effects in the scientific model, the relationship between one of the main effects and the outcome would vary with the other main effect. However, the performance of *improved passive imputation* still remains an open topic [21,27,29]. Finally, if one of the variables involved in the interaction is fully observed, one can impute the data separately in different levels of this variable to allow for the interaction [27]. This has been shown to be less biased than any of the aforementioned methods, but can only be applied in the specific situation in which one of the variables involved in the interaction is fully observed, and is therefore not considered in this paper. The differences in how the imputation model is specified across the *active, passive* and *improved passive* approaches are tabulated in Table 1. Note that all methods impute the auxiliary term, Z, in the same way but that only the *active* approach imputes the interaction term, whereas the *passive* approaches derive it. The main difference across the three approaches lies in how the main effects are imputed.

Several authors have compared these approaches with mixed findings [21,27,29]. von Hippel recommends the *transform–impute* or *active* approach [27]. In his study, he showed that *active imputation* under FCS produced good regression estimates despite the internal inconsistency between the interaction effect and the main effects. He examined two scenarios under MCAR; one in which there was a true interaction between two continuous variables and the other in which there was a true interaction between a continuous and a binary variable. Under the FCS setting, White et al. compared *active, passive*, and *improved passive* approaches when an interaction effect of two binary

Table 2
Difference between scientific and imputation models by imputation technique when interaction term has 2 and 3 levels.

Number of categories in the interaction term	Scientific model	Imputation technique	Imputation model
2	$Y = X_1 + X_2 + X_1X_2$	JM	Y, X_1, X_2, X_1X_2, Z
		FCS	Y, X_1, X_2, X_1X_2, Z
3	$Y = X_1 + X_{2.1} + X_{2.2} + X_1X_{2.1} + X_1X_{2.2}$	JM	$Y, X_1, X_{2.1}, X_{2.2}, X_1X_{2.1}, X_1X_{2.2}, Z$
		FCS	Y, X_1, X_2, X_1X_2, Z

variables are of interest. They found that results from *improved passive* to be lower in bias than that from *passive imputation*. They also showed that *active imputation* led to bias when data are MAR but not MCAR [29]. Seaman et al. conducted a simulation study that yielded more nuanced findings. They compared *active, passive*, and *improved passive* approaches under both MAR and MCAR conditions for continuous and binary outcomes when interaction effects of two continuous variables were of interest. They demonstrated that although *improved passive* was superior to *passive imputation* for reducing bias, *active imputation* performed better than both *passive* approaches when the scientific model was a linear regression under an MAR setting. Under logistic regression, *active imputation* of quadratic terms (not interactions) were evaluated and showed to perform poorly when the outcome was rare. The authors recommend *active imputation* as the best of a set of imperfect imputation methods when estimating interactions using linear regression [21]. In contrast to these findings, van Buuren recommends the use of *passive imputation* to maintain the internal consistency between the interaction effect and the main effects [25].

In particular, imputing multi-level categorical interaction terms warrants attention. When an interaction effect between two nominal categorical variables is of interest (i.e., when the product of two predictors also results in a multi-level categorical term), it is represented in the scientific model by multiple indicator terms. These multiple indicator terms are highly correlated because they are defined to be mutually exclusive. This is distinct from cases in which the interaction is from two continuous variables or two binary variables that result in one term representing the interaction. Even when not involved in interaction effects, nominal categorical variables require careful consideration during the imputation step [1–3,9]. In general, multi-level categorical terms are treated differently under JM and FCS approaches (Table 2). Under JM, multiple indicator variables are included and jointly modeled with all other variables in one imputation model. Under FCS, the multiple indicator variables for the interaction term are modeled together – for example, in a polytomous regression – but separately from all other variables. None of the previous studies discussed earlier investigated situations where the interaction variable is a function of two nominal categorical variables, and no research has assessed the properties resulting from these methods for categorical interaction terms in a linear or a logistic regression model. In this paper we compare and contrast the performance of *active* and *passive imputation* on imputing categorical interaction terms under FCS, and compare the performance of *active imputation* on imputing nominal categorical interaction terms under both JM and FCS. Situations in which the outcome variable is continuous (linear regression analysis) and binary (logistic regression analysis) are considered.

The specific goals of our paper are to (1) Characterize differences in properties of estimates resulting from JM and FCS imputation using an *active* approach, (2) Characterize differences in properties of estimates between *active* and *passive* approaches under FCS, and (3) Provide recommendations and describe the various MI options, specifically when the objective is to estimate the effect of multi-level categorical interaction terms on either a continuous or a binary outcome.

1.1. *Active imputation vs. passive and improved passive imputation under FCS*

When the scientific model involves derived variables such as interaction effects, the user has an option to choose between *active* or *passive imputation*. For example, suppose that the covariates in the scientific model include two partially observed binary variables, X_1 and X_2, and its interaction, X_1X_2, i.e. $Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2$, where β_0 is the intercept and $\beta_{i,\, i=1,2,3}$ are the corresponding coefficients of the terms. In *active imputation*, the interaction variable is imputed as just another variable in the imputation model. This may result in an interaction value, X_1X_2, that is inconsistent with the values of its corresponding main effects (X_1 and X_2). For example, the imputed value of X_1X_2 can be 1 when the imputed values of X_1 and X_2 are both 0. In *active imputation* under FCS, the interaction variable acts like any other variable. As such, it is used to impute all other variables with missing values including its corresponding main effects.

In *passive imputation*, a consistent relationship between X_1, X_2 and X_1X_2 is preserved by deriving X_1X_2 after X_1 and X_2 are imputed. This is only possible under FCS, as FCS is executed on a variable-by-variable basis, where an imputation model is specified for each partially observed variable (whereas JM specifies one imputation model for all variables). As such, users have the flexibility of selecting predictors in the imputation model for each missing variable. When *passively* imputing, X_1X_2 is included in the imputation model for other partially observed variables that are not its corresponding main effects, but X_1X_2 itself is not imputed. Rather, X_1X_2 is derived after X_1 and X_2 are imputed. Importantly, while X_1X_2 is used to impute other missing variables, it is not used to impute X_1 or X_2 (Table 1). This is performed in an iterative process within the imputation procedure until convergence criteria are satisfied. In other words, the interaction variable (X_1X_2) is updated every time one of the two constituent variables (X_1 or X_2) is imputed.

Using the same ideas on which *passive imputation* is based, *improved passive* additionally includes the interaction of a main effect and the outcome as one of the predictors of the other main effect in the imputation model [27,29] (Table 1). More specifically, when imputing X_1, the interaction between Y and X_2 (YX_2) is included as a predictor in the imputation model. Similarly, when imputing X_2, the interaction between Y and X_1 (YX_1) is included.

1.1.1. Imputation model vs. scientific model for FCS

In FCS, imputation models need to be specified for each partially observed variable using Y, X_1, X_2, X_1X_2, and other auxiliary term, Z [15,19]. A binary variable will be imputed using logistic regression methods yielding a variable with binary values that can be directly used in the scientific model. When the variable has more than two categories, polytomous regression or multinomial logistic regression can be used to impute the variable. Because multiple models are being specified (one for each partially observed variable), there is more flexibility than under JM, where one imputation model is specified to jointly impute all variables. This flexibility allows for approaches like *passive* and *improved passive*, where different sets of predictors can be included across each imputation model, or as in the case of *improved passive*, where different functional forms can be considered across each model (Table 1).

1.1.2. R—MICE for FCS

When comparing *active* and *passive imputation* under FCS, we utilized the MICE package in R.

The MICE package, released in 2000, includes procedures to implement FCS for MI in R [26]. Typically, users impute the data with the MICE function, then build m scientific models via functions such as glm, and summarize the estimates using the pool function. The imputation technique for each variable can be specified by the method argument within MICE. Three method options exist for categorical variables; logreg that uses logistic regression, polyreg that uses polytomous (unordered) regression, and lda that uses linear discriminant analysis. The defaults for binary and multi-level categorical variables are logreg and polyreg respectively. In order to use one of these methods, the variable needs to be specified as a factor variable which can be achieved easily using the as.factor() function. The MICE function automatically creates indicator variables for each factor variable that is also being used as a predictor in the imputation model. *Active imputation* is straightforward to perform in MICE; the user simply includes the interaction term just like any other variables in the imputation model and specifies which model to use for imputation. Users of MICE can also perform *passive imputation* by performing a dry run of the imputation and changing the predictor and the method matrices within the program. This requires somewhat intricate coding but is explained very clearly by van Buuren [26]. (Also see example code in Appendix Example Code A.)

1.2. FCS vs. JM using active imputation

Because FCS is executed on a variable-by-variable basis, users have the flexibility of selecting predictors for each missing variable. In *active imputation* under FCS, however, the interaction variable (X_1X_2) acts like any other variable. As such, it is used to impute all other variables with missing values including its corresponding main effects (X_1 and X_2).

Unlike FCS, JM imputes the variables jointly. Therefore, every variable considered in the imputation model is used to impute missing values for all variables, precluding the idea of *passive imputation*. *Active imputation* under JM is comparable to *active imputation* under FCS in its implementation. In both, the interaction effect (X_1X_2) is considered as just another variable and the imputed values do not necessarily preserve the relationship between the interaction and its main effects (X_1 and X_2). In fact, JM active and FCS active are identical when linear regression is used to impute every variable in FCS.

1.2.1. Imputation model vs. scientific model for JM and FCS

Suppose again that the scientific model has the form, $Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2$, where Y is the outcome variable, X_1 and X_2 are partially observed binary variables, and X_1X_2 is the interaction between X_1 and X_2. In both JM and FCS, the imputation model should include the outcome Y, the independent variables of interest, X_1, X_2, X_1X_2, and other auxiliary terms, Z [15,19]. Under JM, the imputed values of X_1 and X_2 will not strictly be binary. In the scientific model, they may be left as continuous or be rounded to produce $\{0, 1\}$ values after the MI procedure. Rounding at 0.5 is highly discouraged as it has been shown to introduce bias [1,2,9], and other rounding approaches that perform better have been proposed [3,4,6,30]. Editing values post-imputation, however, remains a controversial topic. Under FCS, on the other hand, the binary variables are typically imputed using logistic regression methods yielding variables with binary values ready for use in the scientific model. Note that when dealing with binary predictors, the variables specified for inclusion in the imputation models for JM and FCS are equivalent (Table 2).

As described previously, the imputation models differ between JM and FCS when the variable has more than two categories (Table 2). A nominal categorical variable of n levels are expressed as $n - 1$ indicator variables in the scientific model regardless of the imputation approach. In JM, the nominal categorical variable is included in the imputation model in the form of $n - 1$ indicator variables, as expressed in the scientific model. As in the binary case, after imputation, each indicator variable will most likely contain values other than 0 or 1. And, as in the binary case, rounding techniques for multi-level categorical variables have been proposed by some authors to address how the variables are handled in analysis, but this remains an open research topic [3,4,6,30]. Alternatively, the imputed values may be left unedited [1,2,9]. In contrast, the FCS approach imputes multi-level categorical variables using techniques such as polytomous regression or discriminant analysis, where imputed values are ready for analysis without the need for editing or rounding.

1.2.2. SAS—PROC MI for JM

We utilized SAS procedures to implement JM imputation. As with MICE, *active imputation* is straightforward to perform in SAS using PROC MI (PROC MI does not have an option for *passive imputation*). After the MI step, the scientific models from m imputed datasets are fit by a standard model building command such as PROC GENMOD, PROC LOGISTIC, or PROC PHREG, and the parameter estimates are summarized using PROC MIANALYZE.

2. Methods

2.1. Motivating case study

Our work was motivated by a previously published study investigating choices in breast cancer care [13]. This dataset was created from electronic medical records of two independent institutions: an academic medical center and a large multi-site community practice with data sources from institutional and state cancer registries. Before statewide cancer registry data were available (from the California Cancer Registry, a component of the Surveillance, Epidemiology and End Results Program), the investigators only had information from the regional cancer registry, with 41.5% of the cohort missing data on a key variable (Stage: 25.0% missing; Grade: 33.6% missing; Race: 12.4% missing; Ethnicity: 13.6% missing;). The integration of state cancer registry greatly augmented data completion, reducing the missing proportion to 27.4% (Stage: 8.3% missing; Grade: 22.1% missing; Race: 1.9% missing, Ethnicity: 2.0% missing). Because a substantial number of observations were still missing at least one variable, MI-based analysis was considered.

Investigators were particularly interested in differences in patient characteristics by breast cancer treatment procedures, including use of mastectomy, chemotherapy, magnetic resonance imaging (MRI), positron emission tomography (PET), and genetic testing, among patients treated at academic, community or both institutions. Predictors of interest were age at diagnosis, year of diagnosis, race, ethnicity, stage, grade, histology, enrollment in a clinical trial, and refusal of treatment. In addition, synergistic effects among age, year, race and stage on the choice of treatment were of interest. Some of the variables such as stage and grade have an ordinal nature but the study treated these variables as nominal because the investigators were interested in the non-linear effect of these variables on the outcome. Thus, stage was a 5-level (Stages 0–IV) nominal categorical variable and grade was a 3-level (Grades 1–3) nominal categorical variable. In addition, race had four categories (White, Black, Asian, Other), ethnicity had two categories (Hispanic, Non-Hispanic), and patient affiliated institution had three categories (academic, community, both).

2.2. Design of simulation study

We simulated data to resemble the relationships among some of the variables from the breast cancer data. Five scenarios were considered. In each scenario we varied the number of categories in the interaction effect. An interaction effect between; (1) two binary variables, (2) one binary and a 3-level categorical variable, (3) two 3-level categorical variables, (4) one binary and 5-level categorical variable, (5) a 5-level categorical variable and a 3-level categorical variable were considered for Scenarios 1 through 5, respectively. For each scenario, 1000 datasets each with $n = 5000$ observations were simulated. For each of the 5 scenarios, we had an outcome that was continuous and an outcome that was binary. Each dataset included (1) an outcome variable Y (reflecting receipt of mastectomy), (2) two independent nominal categorical variables X_1 (reflecting stage) and X_2 (reflecting patient affiliated institution), their interaction, X_1X_2, and an auxiliary continuous variable, Z (reflecting age). We modeled Z to be related to both X_1 and X_2 such that;

$$Z = 55 - 4X_1 - 3X_2 \text{ (Scenario 1)}$$

$$Z = 57 - 3X_1 + 2X_{21} - 1.5X_{22} \text{ (Scenario 2)}$$

$$Z = 54.5 + 0.1X_{1.1} + 0.4X_{1.2} + 6.1X_{2.1} + 0.5X_{2.2} \text{ (Scenario 3)}$$

$$Z = 54 + 5.6X_1 + 2.6X_{2.1} - 0.7X_{2.2} - 1.9X_{2.3} + 1.0X_{2.4} + 0.4X_1X_{2.1} \text{ (Scenario 4)}$$

$$Z = 54 + 2.6X_{1.1} - 0.7X_{1.2} - 1.9X_{1.3} + 1.0X_{1.4} + 5.6X_{2.1} + 0.4X_{2.2} \text{ (Scenario 5)}$$

where $X_{1.1} = 1$ if $X_1 = 1$ and $X_{1.1} = 0$ otherwise, and similarly for $X_{1.2}, \ldots, X_{2.2}$.

The binary outcomes were generated from a logistic regression model involving X_1, X_2, and X_1X_2 (continuous outcomes used the same linear combination expressed below) such that;

$$\operatorname{logit}(P(Y = 1)) = -1.3 + 1.2X_1 + 0.7X_2 - 0.6X_1X_2 \text{ (Scenario 1)}$$

$$\operatorname{logit}(P(Y = 1)) = -0.5 - 0.2X_1 - 0.1X_{2.1} + 1X_{2.2} + 0.1X_1X_{2.1} + 0.2X_1X_{2.2} \text{ (Scenario 2)}$$

$$\begin{aligned}\operatorname{logit}(P(Y = 1)) = {} & 0.2 + 0.5X_{1.1} + 0.1X_{1.2} - 0.2X_{2.1} - 0.5X_{2.2} - 0.5X_{1.1}X_{2.1} \\ & + 0.2X_{1.1}X_{2.2} + 0.15X_{1.2}X_{2.1} + 0.3X_{1.2}X_{2.2} \text{ (Scenario 3)}\end{aligned}$$

$$\begin{aligned}\operatorname{logit}(P(Y = 1)) = {} & -0.7 - 1.0X_1 + 0.9X_{2.1} + 0.7X_{2.2} + 2.0X_{2.3} + 1.7X_{2.4} - 1.0X_1X_{2.1} \\ & - 1.2X_1X_{2.2} + 1.3X_1X_{2.3} - 1.3X_1X_{2.4} \text{ (Scenario 4)}\end{aligned}$$

$$\begin{aligned}\operatorname{logit}(P(Y = 1)) = {} & -0.7 + 0.1X_{1.1} + 0.7X_{1.2} + 2.0X_{1.3} + 1.5X_{1.4} - 0.3X_{2.1} + 1.0X_{2.2} \\ & - 1.4X_{1.1}X_{2.1} - 1.2X_{1.2}X_{2.1} + 1.3X_{1.3}X_{2.1} - 1.3X_{1.4}X_{2.1} \\ & + 1.4X_{1.1}X_{2.2} - 1.4X_{1.2}X_{2.2} - 1.3X_{1.3}X_{2.2} - 1.0X_{1.4}X_{2.2} \text{ (Scenario 5).}\end{aligned}$$

Missingness of X_1 and X_2 were generated under an MAR condition. The logit of probability of missing for each $X_{i,\, i=1,2}$ was generated as a linear combination of Z such that;

$$\text{logit}(P(X_1 \text{ is missing})) = -7 + 0.12Z \text{ and } \text{logit}(P(X_2 \text{ is missing})) = -1 - 0.02Z \text{ (Scenario 1)}$$

$$\text{logit}(P(X_1 \text{ is missing})) = -5 + 0.05Z \text{ and } \text{logit}(P(X_2 \text{ is missing})) = 0.01 - 0.04Z \text{ (Scenario 2)}$$

$$\text{logit}(P(X_1 \text{ is missing})) = -2.0 + 0.02Z \text{ and } \text{logit}(P(X_2 \text{ is missing})) = -0.65 - 0.015Z \text{ (Scenarios 3–5)}.$$

Both Y and Z were fully observed. The proportion of subjects missing at least one value was set to 40%. In additional simulations, we examined cases in which 20% of subjects were missing at least one variable (Appendix Table A1).

We considered three imputation approaches under FCS using R's MICE package: *active* (*FCS ACTIVE*), *passive* (*FCS PASSIVE*), and *improved passive* (*FCS IMP PASSIVE*). Logistic regression (logreg option) was used to impute binary variables and multinomial regression or polytomous regression (polyreg option) was used to impute categorical variables.

When comparing JM versus FCS techniques under *active imputation*, we considered three methods: *active imputation* under FCS (*FCS ACTIVE*), *active imputation* with no rounding under JM (*JM ACTIVE*), and *active imputation* with rounding under JM (*JMR ACTIVE*). More specifically, for binary variables, imputed values <0.5 were rounded to 0, and values ≥ 0.5 were rounded to 1, and for higher order terms, the rounding method described by Allison was employed [3]. To describe this method briefly, suppose $n - 1$ indicator variables are included in a model to represent a categorical variable with n levels. After calculating the value for the reference category as 1 minus the sum of the $n - 1$ imputed values, the category with the highest value is assigned a value of 1 and the remaining categories are assigned values of 0.

All analyses were based on $m = 10$ imputations. To evaluate methods, we computed the mean point estimate, the bias (defined as the difference between the true parameter value and the estimated value averaged over the simulations), coverage percentage, and mean squared error (MSE). We also computed the jackknife estimate of the Monte Carlo estimate (MCE) of the MSE as a measure of variation within each simulation [12]. For simplicity, we focused on properties of the interaction effects only. JM methods were implemented using SAS's PROC MI and FCS methods were implemented using R's MICE package.

2.3. Illustrative example using real data from a breast cancer study

To illustrate, we applied each method to the breast cancer data. We used both sets of data to examine the impact of each MI method with different proportions of missing data. In our example, we investigated associations between patient characteristics and use of mastectomy. One interaction effect of particular interest (specified *a priori*) was that of patient institutional affiliation and stage. Variables included in the imputation model in addition to the outcome were age, year of diagnosis, stage, grade, patient institutional affiliation, use of PET, enrollment in clinical trial, receipt of genetic testing, and an interaction effect between stage and patient institutional affiliation.

3. Results

3.1. Results from simulation study

3.1.1. Active imputation under FCS vs. JM

Fig. 1 shows the MSEs corresponding to each parameter for the interaction terms across five scenarios when the outcome is continuous for all methods considered. Tables 3a and 3b include average biases and coverage probabilities for each parameter from the simulation in addition to MSEs, for continuous and binary outcomes respectively. We first focused our attention on the three *ACTIVE* methods under continuous outcomes—*active imputation* using FCS (*FCS ACTIVE*), JM (*JM ACTIVE*) and JM with rounding (*JMR ACTIVE*). *FCS ACTIVE* and *JM ACTIVE* performed comparably under Scenarios 1–3, while *JMR ACTIVE* performed worse than the other two. For example, in Scenario 1, where

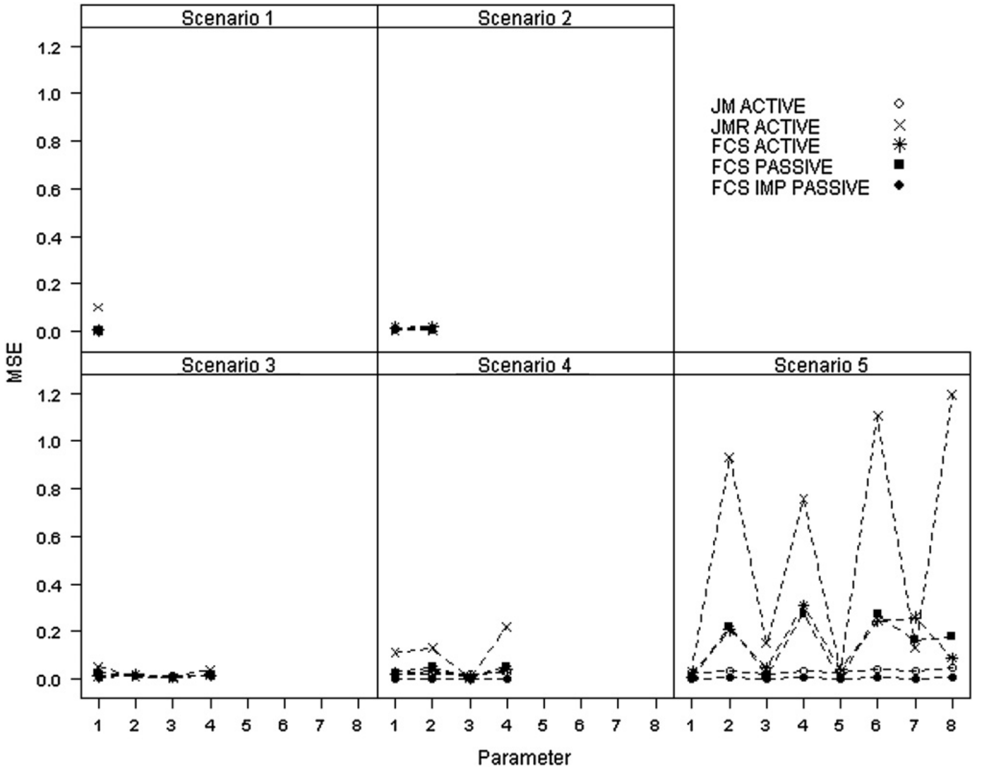

Fig. 1. Mean squared errors (MSE) for interaction parameters estimated using each method for Scenarios 1–5 when outcome is continuous.

there was only one term representing the interaction, the MSE for FCS and JM was 0.007 and 0.006, respectively (Table 3a). For *JMR ACTIVE*, on the other hand, the MSE was 0.103. Variability across the three methods increased for Scenarios 4 and 5 where the number of parameters for the interaction term was 4 and 8, respectively (Fig. 1). *FCS ACTIVE* and *JM ACTIVE* performed comparably in Scenario 4 and *JMR ACTIVE* fared considerably worse with more variable performance across the parameters. In Scenario 5, *JM ACTIVE* performed relatively consistently across the eight parameters. On the other hand, we observed variability in MSEs across the parameters under *FCS ACTIVE* and *JMR ACTIVE*. Specifically, the MSEs for the parameters ranged from 0.010 to 0.310 for *FCS ACTIVE* and from 0.024 to 0.046 for *JM ACTIVE*, whereas for *JMR ACTIVE*, the MSEs were higher with a larger variability ranging from 0.031 to 1.194 (Table 3a). Similar performance was observed with respect to coverage, where coverage was consistent for *JM ACTIVE*, more variable for *FCS ACTIVE*, and poor for *JMR ACTIVE*. For example, in Scenario 5, coverage ranged from 92% to 95% for the eight parameters for *JM ACTIVE* with at least 90% or greater coverage for all parameters. *FCS ACTIVE* yielded a range in coverage probabilities of 3%–96% with at least 90% coverage for only one parameter, and *JMR ACTIVE* gave a range of 0%–92% where three parameters had 0% coverage and two had coverage that was at least 90%. These trends were similar for the binary outcome for both MSE and coverage (Table 3b). To summarize, *JM ACTIVE* and *FCS ACTIVE* gave similar MSEs in Scenarios 1–4, and *JM ACTIVE* was the favored approach in Scenario 5, where it yielded a more stable and consistently lower MSEs across the eight parameters.

3.1.2. Active vs. passive imputation under FCS

Under FCS when the outcome was continuous, *active* (*FCS ACTIVE*) and *passive imputation* (*FCS PASSIVE*) yielded similar MSEs across parameters for all five scenarios (Fig. 1, Table 3a). For Scenarios 1–3, all three methods – *FCS ACTIVE*, *FCS PASSIVE* and *FCS IMP PASSIVE* – were comparable in terms of

Table 3a
Simulation results for each scenario by active imputation method for continuous outcomes.

Scenario	Interaction	True value	JM ACTIVE			*JMR ACTIVE*			*FCS ACTIVE*			*FCS PASSIVE*			*FCS IMP PASSIVE*		
			Bias	MSE	Cov p	Bias	MSE	Cov p	Bias	MSE	Cov p	Bias	MSE	Cov p	Bias	MSE	Cov p
1	*1*	*−0.6*	0.004	0.007	94.7	0.317	0.103	0	0.046	0.006	95.8	0.046	0.006	90.1	0.014	0.005	93.8
2	*1*	*0.1*	0.004	0.012	94.5	−0.041	0.005	97.8	−0.109	0.018	81.4	−0.026	0.007	96.8	−0.016	0.009	91.5
	2	*0.2*	0.002	0.012	91.8	0.039	0.005	96.9	−0.115	0.019	75.9	−0.037	0.007	95.0	−0.012	0.008	92.9
3	*1*	*−0.5*	−0.004	0.012	94.0	0.222	0.053	20.2	0.070	0.011	93.1	0.142	0.025	68.5	−0.004	0.009	93.2
	2	*0.2*	−0.002	0.014	94.9	−0.053	0.008	97.8	−0.117	0.020	89.5	−0.077	0.011	96.8	−0.007	0.011	94.7
	3	*0.15*	−0.001	0.011	95.0	−0.034	0.005	97.6	−0.036	0.007	96.1	−0.054	0.007	95.5	−0.003	0.009	93.5
	4	*0.3*	0.000	0.015	94.8	−0.188	0.041	54.4	−0.093	0.016	93.1	−0.118	0.019	90.5	−0.016	0.013	93.4
4	*1*	*−1*	0.003	0.019	95.3	0.322	0.109	10.3	0.146	0.021	78.8	0.156	0.024	75.6	−0.002	2.41E−06	96.1
	2	*−1.2*	0.002	0.018	94.1	0.354	0.131	2.9	0.189	0.036	63.9	0.221	0.049	49.6	0.001	1.14E−06	95.0
	3	*1.3*	−0.003	0.020	95.1	0.004	0.006	99.7	0.071	0.005	95.0	0.072	0.005	94.0	−0.023	5.24E−04	94.6
	4	*−1.3*	0.003	0.024	94.0	0.463	0.221	0.9	0.196	0.039	73.5	0.228	0.052	60.2	−0.001	1.84E−06	94.8
5	*1*	*−1.4*	0.006	0.026	94.7	0.156	0.031	91.7	0.098	0.010	96.2	0.106	0.011	94.1	0.017	2.73E−04	93.0
	2	*−1.2*	−0.001	0.037	92.8	0.962	0.933	0	0.455	0.207	16.8	0.469	0.220	14.1	0.066	0.004	92.0
	3	*1.3*	0.000	0.024	94.6	−0.381	0.152	8.2	−0.215	0.046	76.7	−0.104	0.011	95.4	0.015	2.38E−04	93.7
	4	*−1.3*	0.003	0.035	93.1	0.866	0.756	0	0.557	0.310	3.2	0.524	0.274	5.4	0.069	0.005	91.5
	5	*1.4*	0.005	0.029	93.9	−0.167	0.036	90.0	−0.198	0.039	83.3	0.043	0.002	97.7	0.024	0.001	93.9
	6	*−1.4*	0.005	0.041	92.2	1.046	1.104	0	0.493	0.243	13.0	0.520	0.270	9.2	0.081	0.007	90.3
	7	*−1.3*	0.001	0.034	94.1	0.347	0.131	42.9	0.507	0.257	26.2	0.406	0.165	28.5	0.017	2.77E−04	92.3
	8	*−1*	0.007	0.046	93.2	1.087	1.194	0	0.297	0.088	72.4	0.422	0.178	36.9	0.062	0.004	92.0

Cov p—Coverage probability, MSE—Mean squared error.

Table 3b
Simulation results for each scenario by active imputation method for binary outcomes.

Scenario	Interaction	True value	*JM ACTIVE*			*JMR ACTIVE*			*FCS ACTIVE*			*FCS PASSIVE*			*FCS IMP PASSIVE*		
			Bias	MSE	Cov p	Bias	MSE	Cov p	Bias	MSE	Cov p	Bias	MSE	Cov p	Bias	MSE	Cov p
1	*1*	*−0.6*	0.046	0.036	92.3	0.349	0.132	20.8	0.044	0.022	96.3	0.042	0.020	94.4	0.022	0.022	94.3
2	*1*	*0.1*	0.010	0.048	95.5	−0.036	0.015	98.9	−0.139	0.041	93.3	−0.030	0.022	98.1	0.017	0.035	94.6
	2	*0.2*	0.005	0.047	94.3	0.045	0.016	98.5	−0.150	0.044	90.6	−0.043	0.022	97.8	−0.028	0.034	94.4
3	*1*	*−0.5*	0.009	0.049	94.5	0.231	0.069	79.3	0.073	0.028	96.6	0.149	0.040	94.4	−0.004	0.037	94.4
	2	*0.2*	0.009	0.066	93.1	−0.050	0.024	98.5	−0.129	0.047	96.7	−0.083	0.029	99.4	−0.022	0.051	94.3
	3	*0.15*	−0.006	0.047	94.3	−0.036	0.017	98.6	−0.039	0.023	98.3	−0.056	0.021	98.5	−0.017	0.037	93.8
	4	*0.3*	0.007	0.061	94.2	−0.186	0.054	91.8	−0.098	0.039	97.1	−0.119	0.036	98.2	−0.039	0.051	94.3
4	*1*	*−1*	−0.102	0.125	93.2	0.347	0.147	80.1	0.112	0.013	96.1	0.106	0.011	96.4	−0.008	5.72E−05	94.3
	2	*−1.2*	0.000	0.108	94.3	0.510	0.284	41.1	0.304	0.093	84.2	0.258	0.067	88.8	−0.011	1.15E−04	93.9
	3	*1.3*	0.027	0.128	95.9	−0.117	0.044	98.0	−0.088	0.008	96.5	−0.120	0.014	95.5	0.018	3.08E−04	95.2
	4	*−1.3*	−0.201	0.183	91.5	0.484	0.272	63.6	0.186	0.035	95.2	0.193	0.037	95.2	0.002	3.34E−06	95.3
5	*1*	*−1.4*	0.253	0.196	90.7	0.415	0.196	73.7	0.132	0.018	98.1	0.020	3.89E−04	98.9	0.015	2.10E−04	93.9
	2	*−1.2*	−0.096	0.173	94.1	0.934	0.900	1.2	0.452	0.204	79.0	0.484	0.235	77.9	0.105	0.011	94.8
	3	*1.3*	0.097	0.129	94.3	−0.461	0.232	50.7	−0.587	0.345	41.4	−0.245	0.060	94.3	0.044	0.002	94.0
	4	*−1.3*	−0.109	0.169	92.5	0.807	0.675	2.5	0.555	0.308	59.0	0.524	0.275	69.6	0.128	0.016	93.0
	5	*1.4*	−0.230	0.208	91.5	−0.439	0.224	78.5	−0.600	0.360	64.3	−0.178	0.032	96.5	0.054	0.003	93.4
	6	*−1.4*	−0.066	0.195	93.9	1.046	1.135	1.7	0.458	0.210	82.5	0.532	0.283	77.8	0.169	0.029	93.7
	7	*−1.3*	−0.103	0.175	93.3	0.346	0.156	86.6	−0.090	0.008	98.4	0.249	0.062	94.6	0.028	0.001	93.6
	8	*−1*	−0.076	0.216	94.0	1.080	1.219	2.9	0.285	0.081	94.8	0.425	0.181	90.5	0.098	0.010	95.5

Cov p—Coverage probability, MSE—Mean squared error.

Table 4a
Effect of stage and patient affiliated institution on mastectomy by imputation method (27.4% missing).

	Complete case ($N = 9093$)	MI analysis ($N = 12, 115$)				
		JM ACTIVE	*JMR ACTIVE*	*FCS ACTIVE*	*FCS PASSIVE*	*FCS IMP PASSIVE*
Stage 0						
Academic	1.00	1.00	1.00	1.00	1.00	1.00
Community	0.99 (0.77, 1.27)	0.83 (0.69, 1.00)	0.77 (0.64, 0.92)	0.72 (0.59, 0.88)	0.73 (0.59, 0.89)	0.73 (0.59, 0.91)
Both	2.41 (1.86, 3.12)	2.09 (1.67, 2.63)	1.91 (1.54, 2.38)	1.99 (1.60, 2.49)	2.33 (1.83, 2.96)	2.31 (1.82, 2.92)
Stage I						
Academic	1.00	1.00	1.00	1.00	1.00	1.00
Community	0.79 (0.67, 0.94)	0.63 (0.53, 0.74)	0.63 (0.54, 0.74)	0.71 (0.55, 0.91)	0.71 (0.54, 0.91)	0.70 (0.54, 0.92)
Both	1.79 (1.48, 2.16)	1.85 (1.54, 2.22)	1.81 (1.49, 2.19)	1.88 (1.42, 2.50)	1.81 (1.34, 2.44)	1.82 (1.35, 2.45)
Stage II						
Academic	1.00	1.00	1.00	1.00	1.00	1.00
Community	1.20 (1.01, 1.43)	0.89 (0.76, 1.04)	0.90 (0.76, 1.05)	0.95 (0.74, 1.22)	0.89 (0.69, 1.15)	0.89 (0.68, 1.16)
Both	1.76 (1.45, 2.13)	1.95 (1.63, 2.33)	1.96 (1.63, 2.35)	2.00 (1.50, 2.66)	1.87 (1.38, 2.53)	1.90 (1.41, 2.55)
Stage III						
Academic	1.00	1.00	1.00	1.00	1.00	1.00
Community	1.46 (0.95, 2.23)	1.07 (0.76, 1.51)	1.27 (0.89, 1.80)	0.80 (0.54, 1.19)	0.90 (0.61, 1.33)	0.95 (0.64, 1.40)
Both	1.70 (1.06, 2.74)	1.65 (1.12, 2.44)	2.02 (1.35, 3.03)	1.62 (0.99, 2.65)	1.70 (1.06, 2.71)	1.61 (1.01, 2.56)
Stage IV						
Academic	1.00	1.00	1.00	1.00	1.00	1.00
Community	0.50 (0.23, 1.08)	0.39 (0.18, 0.88)	0.54 (0.27, 1.09)	0.45 (0.22, 0.91)	0.43 (0.21, 0.91)	0.41 (0.20, 0.86)
Both	0.79 (0.37, 1.71)	0.55 (0.29, 1.06)	0.71 (0.38, 1.36)	0.63 (0.32, 1.25)	0.61 (0.31, 1.21)	0.58 (0.29, 1.14)

MSEs. For Scenario 4, however, *FCS ACTIVE* and *FCS PASSIVE* presented MSEs that were slightly higher and more variable than those presented by *FCS IMP PASSIVE*. In Scenario 5, while *FCS ACTIVE* and *FCS PASSIVE* were comparable, both provided MSEs that were considerably worse than that provided by *FCS IMP PASSIVE*, which yielded relatively low MSEs across all eight parameters. In general, coverage probabilities obtained from *FCS IMP PASSIVE* were superior to the other two methods. For example, in Scenario 5, all coverage probabilities were over 90%, ranging from 90% to 94%, whereas coverage probabilities for *FCS PASSIVE* ranged from 5% to 98%, with only three of the eight parameters having coverage probabilities over 90%. Similarly, *FCS ACTIVE* had a range of coverage probabilities from 3% to 96%, with only one parameter having coverage over 90%. Similar trends were observed when the outcome was binary (Table 3b). In summary, *FCS IMP PASSIVE* had superior performance over the other two FCS methods. Performances between *FCS ACTIVE* and *FCS PASSIVE* were fairly comparable across all scenarios.

3.1.3. JM active versus FCS improved passive

Although when using an *active* approach, JM was favored over FCS, we had evidence that *improved passive* under FCS was superior to *active* under FCS. This left the question of whether *improved passive* under FCS was also superior to *active* under JM. Directly comparing *JM ACTIVE* versus *FCS IMP PASSIVE* across the 5 scenarios showed that *FCS IMP PASSIVE* was favored over *JM ACTIVE* for both continuous and binary outcomes. For the continuous outcome, the two approaches were comparable for Scenarios 1–4. When the number of parameters representing the interaction was large as in Scenario 5, *FCS IMP PASSIVE* outperformed *JM ACTIVE* for all parameters. The difference in performance was larger, however, when the outcome was binary. Improvement in *FCS IMP PASSIVE* was apparent when the number of parameters was four or greater as in Scenarios 3–5. The average MCE estimates of the MSE across the parameters were low – 0.0002, 0.0004, 0.0005, 0.0006, 0.0071 – for Scenarios 1 through 5, respectively, with continuous outcome, indicating the number of simulations on which we base our findings was adequate.

3.2. Results from breast cancer study

Prior to the merge with the state cancer registry, the cohort consisted of 8605 women. The proportion of women missing data was 41.5%. After integrating data from the state cancer registry, the number of women in the study cohort increased to 12,115. Of those, 27.4% had at least one missing value. Variables with missing values in the scientific model were race, ethnicity, stage, grade, histology, and the interaction between stage and patient affiliation. Although stage and grade can be considered ordinal categorical variables, we treated them as nominal since we were interested in their non-linear effect on the outcome. Therefore the interaction variable consisted of stage—a 5-level (Stages 0–IV) nominal categorical variable—and patient affiliation—a 3-level (academic, community-based, and both) nominal categorical variable. Table 4a shows the adjusted odds ratios (ORs) and 95% confidence intervals (CIs) corresponding to the interaction effects from the cohort where 27.4% of subjects were missing at least one variable, and Table 4b shows these results for the cohort where 41.5% are missing.

Based on the recently derived cohort with 27% of subjects missing data (Table 4a), the CC analyses suggested that patients affiliated with both institutions were significantly more likely to receive mastectomy than patients affiliated with the academic institution across all stages except for the highest stage category. Among patients with stage I cancer, those affiliated with the community institution were less likely to receive mastectomy compared to patients affiliated with the academic institution, but among patients with stage II cancer, community-affiliated patients were more likely to receive mastectomy. Notable differences were observed between the CC analysis and MI-based analyses in describing associations between affiliation and mastectomy for stage II and stage IV patients. Specifically, the OR (95% CI) indicated a 20% increased use of mastectomy for community-affiliated patients of stage II relative to academic-affiliated patients (OR = 1.20; 95% CI from 1.01 to 1.43) in the CC analysis whereas the MI-based analyses suggested no difference in mastectomy rates between these patients (for example, OR from *FCS ACTIVE* = 0.95; 0.74–1.22, OR from *FCS PASSIVE* = 0.89; 0.69–1.15). Among

Table 4b
Effect of stage and patient affiliated institution on mastectomy by imputation method (41.5% missing).

	Complete case ($N = 5035$)	MICE ($N = 8605$)				
		JM ACTIVE	*JMR ACTIVE*	*FCS ACTIVE*	*FCS PASSIVE*	*FCS IMP PASSIVE*
Stage 0						
Academic	1.00	1.00	1.00	1.00	1.00	1.00
Community	1.22 (0.86, 1.73)	0.93 (0.77, 1.13)	0.99 (0.83, 1.17)	1.28 (1.03, 1.59)	1.06 (0.83, 1.35)	1.04 (0.83, 1.32)
Both	2.52 (1.51, 4.21)	1.55 (1.08, 2.23)	1.14 (0.78, 1.67)	1.43 (0.91, 2.23)	1.80 (1.23, 2.65)	1.96 (1.32, 2.91)
Stage I						
Academic	1.00	1.00	1.00	1.00	1.00	1.00
Community	1.11 (0.90, 1.36)	1.31 (1.07, 1.60)	1.19 (0.99, 1.41)	1.32 (0.98, 1.77)	1.34 (0.99, 1.80)	1.32 (0.96, 1.80)
Both	1.51 (1.10, 2.06)	1.72 (1.29, 2.29)	1.65 (1.24, 2.21)	1.80 (1.04, 3.10)	1.64 (1.01, 2.64)	1.64 (1.01, 2.67)
Stage II						
Academic	1.00	1.00	1.00	1.00	1.00	1.00
Community	1.58 (1.28, 1.95)	1.82 (1.49, 2.23)	1.71 (1.38, 2.11)	1.35 (1.00, 1.82)	1.45 (1.07, 1.96)	1.47 (1.09, 1.97)
Both	1.19 (0.88, 1.61)	1.30 (0.98, 1.72)	1.41 (1.09, 1.83)	1.35 (0.76, 2.41)	1.31 (0.81, 2.12)	1.29 (0.79, 2.10)
Stage III						
Academic	1.00	1.00	1.00	1.00	1.00	1.00
Community	1.82 (1.15, 2.88)	2.30 (1.43, 3.71)	2.63 (1.77, 3.91)	1.49 (0.99, 2.25)	1.37 (0.85, 2.21)	1.60 (1.00, 2.55)
Both	1.40 (0.77, 2.56)	1.70 (1.02, 2.83)	2.49 (1.48, 4.17)	1.85 (0.94, 3.64)	1.63 (0.88, 3.01)	1.64 (0.86, 3.11)
Stage IV						
Academic	1.00	1.00	1.00	1.00	1.00	1.00
Community	0.70 (0.28, 1.71)	0.50 (0.20, 1.24)	0.65 (0.31, 1.36)	0.65 (0.32, 1.31)	0.62 (0.29, 1.32)	0.53 (0.25, 1.09)
Both	0.75 (0.27, 2.13)	0.61 (0.27, 1.35)	0.85 (0.38, 1.91)	0.79 (0.32, 1.95)	0.84 (0.35, 2.01)	0.73 (0.30, 1.79)

stage IV patients, the CC analysis indicated a non-significant reduction of 50% in use of mastectomy for community-affiliated patients (OR = 0.50; 0.23–1.08) whereas the MI-based analyses – with the exception of *JMR ACTIVE* – suggested a significant reduction in use of mastectomy. Specifically, findings from *JM ACTIVE* yielded a 61% reduction in use (OR = 0.39, 95% CI from 0.18 to 0.88) and *FCS IMP PASSIVE* yielded a 59% reduction in use (OR = 0.41, 95% CI from 0.20 to 0.86). Results from the various MI approaches were generally comparable with ORs consistently in the same direction, although moderate variability was observed. Variability across MI approaches was greater in the cohort with 41.5% missing (Table 4b). For example, among Stage 0 patients, the ORs for those affiliated with both academic and community institutions ranged from 1.14 (*JMR ACTIVE*) to 1.96 (FCS *IMP PASSIVE*) where some methods suggested a significant increase in the OR for use (*JM ACTIVE*, *FCS PASSIVE* and *FCS IMP PASSIVE*) but others did not (*JMR ACTIVE*, *FCS ACTIVE*). In addition, among Stage III patients, findings based on *JM ACTIVE*, *JMR ACTIVE*, and *FCS IMP PASSIVE* suggested an increase in use for those affiliated with a community health center whereas *FCS ACTIVE* and *FCS PASSIVE* did not. For example, based on the analysis from *FCS IMP PASSIVE*, Stage III patients affiliated with community health center have a 60% increase in mastectomy relative to those at an academic health center, whereas an analysis based on *FCS PASSIVE* would conclude that there is no such association between affiliation and mastectomy among Stage III patients.

Based on our simulation study, we would interpret our findings from the results generated by *FCS IMP PASSIVE*. The findings indicated an interesting interaction effect between the type of institution and stage on receiving a mastectomy (Table 4a). With the exception of Stage IV, those affiliated with both types of institutions were more likely to receive a mastectomy than those exclusively seen at an academic institution. In addition, Stage I and Stage IV patients who received their care exclusively at a community center were less likely than those at an academic health center to receive a mastectomy.

4. Discussion

Through simulations, we characterized variability between *active* and *passive* options for performing MI in the presence of interaction effects involving nominal categorical terms. In the presence of variability across methods, our goal was to identify the best method for imputing data in the presence of categorical interaction effects.

We observed comparable performance between *active* and *passive imputation* under FCS, and found *improved passive imputation* under FCS to be superior to both *active* and *passive* approaches particularly when the number of parameters was large (Scenarios 4 and 5). This was the case for both continuous and binary outcomes. We also compared JM and FCS techniques under *active imputation*. *Active imputation* under JM was favored over *active imputation* with rounding under JM and over *active imputation* under FCS. As in the comparison between *active* and *passive* approaches, differences among methods were most apparent when the number of parameters was large. Given that *active imputation* under JM was the clear winner, we compared this approach to the *improved passive* approach under FCS. We found the *improved passive* approach to be superior to *active imputation* under JM. Our results from additional simulation studies where the percentage missing was lower and the outcome was binary were comparable to our main results. For all methods, the biases were generally smaller when the proportion of observations with missing was reduced to 20% (see Table A1 in Appendix). Based on these findings we recommend the use of the *improved passive* approach using FCS imputation. The *improved passive* approach successfully incorporates the relationship between the outcome variable and each of the main effects, thus getting closer to the ideal but impractical method of stratifying the imputation by levels of one of the predictors [27,29].

Previously, von Hippel observed *active imputation* to be less biased than *passive imputation* when imputing interaction terms from two binary variables under MCAR setting. We conducted our study under the MAR setting and found that *active* and *passive imputations* were comparable when imputing interaction terms from two binary variables. Seaman et al. [21] evaluated *improved passive* along with *active* and *passive* approaches via simulations under the MAR setting. The authors only examined interaction effects of two continuous predictors on a continuous outcome. Thus, interest centered on the case where there was one term representing the interaction effect, whereas we focused on the case where multiple terms corresponding to the interaction term are of interest. Seaman et al. found that while *improved passive* reduced bias relative to *passive* considerably, *active* produced estimates with the least bias. Our findings, however, are relevant for multi-level categorical interaction terms. When one interaction term was of interest, we found the methods to be comparable under MAR. Whereas we used MSE to gauge performance, Seaman et al. used bias and coverage. Estimates of coverage in Scenario 1 were comparable to what Seaman et al. obtained. For example, Seaman et al. obtained coverage estimates of 92%, 94%, and 94% for *active, passive*, and *improved passive*, respectively when the R^2 statistic corresponding to the model fit was 0.5. In Scenario 1, we obtained estimates of 95.8%, 90.1% and 93.8%. Because the authors do not provide estimates of the standard errors obtained from the models, it is difficult to directly compare the performance of methods in their scenarios to the performance in ours, where we consider both bias and efficiency as important components of judging performance.

Findings from our simulations confirmed results of previous studies that showed rounding imputed values post imputation to create categorical variables introduces bias [1,2,9]. In each scenario, results from *active imputation* with rounding under JM yielded the largest MSE and the lowest coverage probability among all methods we evaluated, expect in Scenario 2. We found that using the unedited imputed values directly performed much better by yielding consistently low MSE and coverage probabilities above 90.0% across all scenarios.

Our study has some limitations. We did not examine every possible method for multiply imputing nominal categorical variables and their interaction. Instead we considered methods that required minimal additional programming, as we wanted our recommendations to apply to those relying on user-friendly packages. For example, alternative rounding methods such as the calibration technique proposed by Yucel for binary and categorical variables may also be applied for JM approaches [30,31]. Demirtas also has proposed rounding strategies for binary variables that utilize information from other variables in the MI model [6]. Because these methods require the user to program additional code, we did not consider them, as our focus is on assessing methods readily available in mainstream software packages. Our study is also limited in the number and types of scenarios we considered. For example, other MAR conditions could be explored (e.g., where missingness is also related to the outcome), and possibly also the NMAR condition. Another scenario we did not consider is one where the auxiliary term, Z, is also missing or one where the outcome variable, Y, is also missing. Whereas we examined a continuous and binary outcome, other non-normal distributions such as the log-normal may yield different findings. Consequently, the credibility of our results is based solely on simulations; our findings could possibly exhibit equivocal conclusions with data generated under a different set of assumptions.

Our simulation study has important strengths. We considered many types of interactions from simple to complicated categorical variables under a flexible and realistic missing data mechanism while varying the proportions of missing data. We examined options available in a mainstream software package that is already fully developed and well used. Other notable programs that perform FCS include mi and ice functions in Stata, and the fairly new FCS option in SAS PROC MI (available version 9.3 and after). See Appendix for more details. As far as we know, this is the first paper to examine *active* and *passive* approaches for multi-level categorical interaction terms. Because multiple correlated parameters are involved in such interaction effects, performing MI-based analysis can pose additional challenges beyond performing MI-based analysis on typical interaction effects that can be represented with one parameter. In addition, FCS and JM treat interaction effects differently when multiple terms are involved. Furthermore, numerous choices under FCS are possible. Our study allows specific recommendations for performing MI in the presence of such complicated interaction terms.

4.1. Recommendations

Based on our study, we recommend implementing *improved passive imputation* under FCS when nominal categorical interactions are to be included in the analysis. Our study highlights one of the barriers of incorporating MI into analyses—that of making additional choices, which can increase the burden on the user. As a general principle, because different choices may produce variable results, we highly recommend the use of sensitivity analyses. This may involve including a variety of different imputation techniques as we presented here (e.g. purely *passive* or *active imputation* under FCS and/or *active imputation* under JM), and/or it may involve inclusion of different sets of auxiliary terms. Presenting a range of results to the readers in the presence of variability may provide insight into the robustness of the findings.

Acknowledgments

This research was supported in part by: The Patient-Centered Outcomes Research Institute Grant ME-1303-5989; The Susan and Richard Levy Gift Fund; Suzanne Pride Bryan Fund for Breast Cancer Research; the Breast Cancer Research Foundation; Regents of the University of California's California Breast Cancer Research Program (16OB-0149 and 19IB-0124); Stanford University Developmental Research Fund; and the National Cancer Institute's Surveillance, Epidemiology and End Results Program under contract HHSN261201000140C awarded to the Cancer Prevention Institute of California. The project was also supported by an NIH CTSA award number UL1 RR025744. The collection of cancer incidence data used in this study was supported by the California Department of Health Services as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885; the National Cancer Institute's Surveillance, Epidemiology, and End Results Program under contract HHSN261201000140C awarded to the Cancer Prevention Institute of California, contract HHSN261201000035C awarded to the University of Southern California, and contract HHSN261201000034C awarded to the Public Health Institute; and the Centers for Disease Control and Prevention's National Program of Cancer Registries, under agreement #1U58 DP000807-01 awarded to the Public Health Institute. The ideas and opinions expressed herein are those of the authors, and endorsement by the University or State of California, the California Department of Health Services, the National Cancer Institute, or the Centers for Disease Control and Prevention or their contractors and subcontractors is not intended nor should be inferred. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the position or policy of the Patient-Centered Outcomes Research Institute, or the United States government.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.stamet.2015.06.001.

References

[1] C. Ake, Rounding after multiple imputation with non-binary categorical covariates, in: SAS Users Group International 30: 2005, 2005.
[2] P. Allison, Imputation of categorical variables with PROC MI, in: SAS Users Group International 30: 2005; 2005.
[3] P. Allison, Missing Data, Sage, Thousand Oaks, CA, 2002.
[4] C.A. Bernaards, T.R. Belin, J.L. Schafer, Robustness of a multivariate normal approximation for imputation of incomplete binary data, Stat. Med. 26 (6) (2007) 1368–1382.
[5] L.M. Collins, J.L. Schafer, C.M. Kam, A comparison of inclusive and restrictive strategies in modern missing data procedures, Psychol. Methods 6 (4) (2001) 330–351.
[6] H. Demirtas, Rounding strategies for multiply imputed binary data, Biom. J. 51 (4) (2009) 677–688.
[7] O. Harel, X.H. Zhou, Multiple imputation: Review of theory, implementation and software, Stat. Med. 26 (16) (2007) 3057–3077.
[8] N.J. Horton, K.P. Kleinman, Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models, Amer. Statist. 61 (1) (2007) 79–90.
[9] N.J. Horton, S.R. Lipsitz, M. Parzen, A potential for bias when rounding in multiple imputation, Amer. Statist. 57 (4) (2003) 229–232.
[10] J.G. Ibrahim, M.H. Chen, S.R. Lipsitz, A.H. Herring, Missing-data methods for generalized linear models: A comparative review, J. Amer. Statist. Assoc. 100 (469) (2005) 332–346.
[11] M.G. Kenward, J. Carpenter, Multiple imputation: current perspectives, Stat. Methods Med. Res. 16 (3) (2007) 199–218.
[12] E. Koehler, E. Brown, S.J. Haneuse, On the assessment of Monte Carlo error in simulation-based statistical analyses, Amer. Statist. 63 (2) (2009) 155–162.
[13] A.W. Kurian, A. Mitani, M. Desai, P.P. Yu, T. Seto, S.C. Weber, C. Olson, P. Kenkare, S.L. Gomez, M.A. de Bruin, et al., Breast cancer treatment across health care systems: Linking electronic medical records and state registry data to enable outcomes research, Cancer 120 (1) (2014) 103–111.
[14] R. Little, D. Rubin, Statistical Analysis with Missing Data, second ed., J. Wiley & Sons, New York, 2002.
[15] X. Meng, Multiple-imputation inferences with uncongenial sources of input, Statist. Sci. 9 (4) (1994) 538–558.

[16] R Core Team R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. In. Vienna, Austria; 2012.
[17] P. Royston, Multiple imputation of missing values: Further update of ice, with an emphasis an categorical variables, Stata J. 9 (3) (2009) 466–477.
[18] D.B. Rubin, Multiple Imputation for Nonresponse in Surveys, J. Wiley & Sons, New York, 1987.
[19] D.B. Rubin, Multiple imputation after 18+ years, J. Amer. Statist. Assoc. 91 (434) (1996) 473–489.
[20] SAS/STAT 9.3 user's guide. In. Cary, NC: SAS Institute, Inc.; 2011.
[21] S.R. Seaman, J.W. Bartlett, I.R. White, Multiple imputation of missing covariates with non-linear effects and interactions: an evaluation of statistical methods, BMC Med. Res. Methodol. 12 (2012).
[22] StataCorp: Stata Statistical Software: Release 12. In. College Station, TX: StataCorp LP; 2011.
[23] J.A.C. Sterne, I.R. White, J.B. Carlin, M. Spratt, P. Royston, M.G. Kenward, A.M. Wood, J.R. Carpenter, Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls, Br. Med. J. 339 (2009) 9.
[24] S. van Buuren, Multiple imputation of discrete and continuous data by fully conditional specification, Stat. Methods Med. Res. 16 (3) (2007) 219–242.
[25] S. van Buuren, Flexible Imputation of Missing Data, Chapman & Hall/CRC Press, Boca Raton, FL, 2012.
[26] S. van Buuren, K. Groothuis-Oudshoorn, mice: Multivariate imputation by chained equations in R, J. Stat. Softw. 45 (3) (2011) 1–67.
[27] P.T. von Hippel, How to impute interactions, squares and other transformed variables, Sociol. Methodol. 39 (2009) 265–291.
[28] J. Ware, D. Harrington, D. Hunter, R. D'Agostino, Missing Data, N. Engl. J. Med. 367 (14) (2012) 1353–1354.
[29] I.R. White, P. Royston, A.M. Wood, Multiple imputation using chained equations: Issues and guidance for practice, Stat. Med. 30 (4) (2011) 377–399.
[30] R.M. Yucel, Y. He, A.M. Zaslavsky, Using calibration to improve rounding in imputation, Amer. Statist. 62 (2) (2008) 125–129.
[31] R.M. Yucel, Y. He, A.M. Zaslavsky, Gaussian-based routines to impute categorical variables in health surveys, Stat. Med. 30 (29) (2011) 3447–3460.