# Real-time handling of missing data in the application of prediction models: a comparison of methods

Nijman SWJa\*, Oberman HIb\*, Brandjes Mc, Jacobs JJLc, Bots MLa, Asselbergs FWdef, Moons KGMa, Vink Gb, Debray TPAaf, Smeden van Ma

a Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands;

b Department of Methodology and Statistics, Utrecht University, Utrecht, the Netherlands

c Department of Health, Ortec B.V. Zoetermeer, The Netherlands;

d Department of Cardiology, University Medical Center Utrecht, Utrecht University, The Netherlands;

e Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, London, United Kingdom;

f Health Data Research UK, Institute of Health Informatics, University College London, London, United Kingdom

\* Equal contribution

Corresponding author: Steven WJ Nijman, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX, Utrecht, the Netherlands. E-mail: s.w.j.nijman@umcutrecht.nl. Telephone: +31-(0)88-75 6801

# Abstract

**Introduction –** The need to solve for missing values in real time is unique to the application of prediction models. The topic of real-time imputation is underrepresented in the literature. In this study, we aim to evaluate various real-time strategies to handle the pervasive problem of missing data when using clinical data. We aim to evaluate the influence of built-in missing data handling mechanisms on prediction accuracy and compare it with existing real-time imputation methods (e.g., joint modeling imputation). We evaluate the effect of various missing data handling methods under specific missing data circumstances as would occur in medical practice.

**Methods –**

**Results –**

**Discussion –**

# Introduction

Incompleteness of medical records is a ubiquitous problem when using healthcare data. Besides the well-documented issues that missing data can create in data analyses, incompleteness of medical records may also create practical issues in clinical practice (1,2). For instance, a prediction model that relies on historical but unrecorded data for a particular patient or prediction models that are used as early-warning systems for individual patients (3,4). Most prediction models are not designed to be used when predictors are not fully observed, and ad-hoc approaches such as replacing the missing value with the population average value (i.e., mean imputation) is generally not advised (1,5). As prediction models are increasingly being integrated in the electronic health record (EHR) via clinical decision support systems (CDSS), the substantial impediment of missing data on the direct use of prediction models in individual patients becomes more evident (6,7). The issue is further compounded as the (gold) standard strategies to mend or circumvent missing data are not suited for use in individual patient data in live clinical practice.

Various strategies to handle challenging manifestations of missing data have been studied thoroughly and can usually provide more plausible substitution values (e.g., via imputation) (2). Multiple imputation is often considered to be the gold standard and can provide valid estimates and correct standard errors when the solution to the problem does not depend on the unobserved values (8). Most imputation algorithms, however, require direct access to data from multiple instances (i.e., multiple patients or multiple measurements) and are therefore not suitable for use on a case-by-case basis. Further, when a prediction model is applied to a single patient in clinical practice via a CDSS there is (usually) no access to any data from other individuals due to computational and privacy constraints [ref].

An intuitive alternative to imputation is to solve for the missingness inside the prediction model instead of the data. Two promising methods of this type are the pattern submodel (PS) approach or surrogate splits (SS). Pattern submodels are attractive to a variety of parameter-based modeling techniques (e.g. regression). The so-called submodels incorporate the nature of the missing data by developing a separate prediction model for all possible missing data patterns (11). Then, when applied to a new case or out-of-sample individual the corresponding prediction model that matches the individual’s missing data pattern is used. Whereas the PS approach lends itself to various kinds of prediction models, the surrogate split approach comes naturally to random forest models (9,10). Briefly, these surrogate splits attempt to preserve the partitioning of the original split by finding the next most optimal split given other observed variables. When the model is applied, each original split for which the predictor is missing will be replaced by the best available ‘surrogate’ variable to decide the split direction (9,10).

In this article we compare various real-time missing data handling approaches when implementing specific modeling techniques in live clinical practice. We use the term 'real-time' to refer to methods that can be applied to data from a single individual as would occur in clinical practice, without necessitating the use of data from other individuals at the point of care. We present an extensive simulation study and a motivating example to compare the different missing data handling strategies that can be used at the implementation level. The aim is to identify strengths and weaknesses of these approaches on the ability to estimate individualized risk, as quantified by the discrimination and calibration of the predictions.

**~~Motivating example~~**

~~Next to our simulation study, we also evaluate the effect of the built-in methods and real-time imputation models when used in actual patients from the large Medical Information Mart for Intensive Care (MIMIC)-III dataset (17). MIMIC-III provides a large database which contains information about patients staying in critical care units of the single tertiary care Beth Israel Deaconess Medical Center.~~

~~Similarly, to the simulation study, the prediction models of interest were the flexible logistic regression and random forest models. We derived both models in MIMIC-III using predictors from existing relevant prediction models using mortality as the primary outcome. For the logistic regression, we considered the Sequential Organ Failure Assessment (SOFA) prediction score and for the random forest model we considered the … (18). The SOFA score estimates the number and severity of failed organs, with in-hospital mortality as the primary outcome.~~

# Missing data handling methods for prediction models

We consider the following three prediction modeling strategies for real-time handling of missing data: (i) prediction models that adopt joint modeling imputation, (ii) prediction models that adopt apattern submodel approach (iii) prediction models that adopt random forests with surrogate splits (10–13).

## Joint Modeling Imputation (JMI)

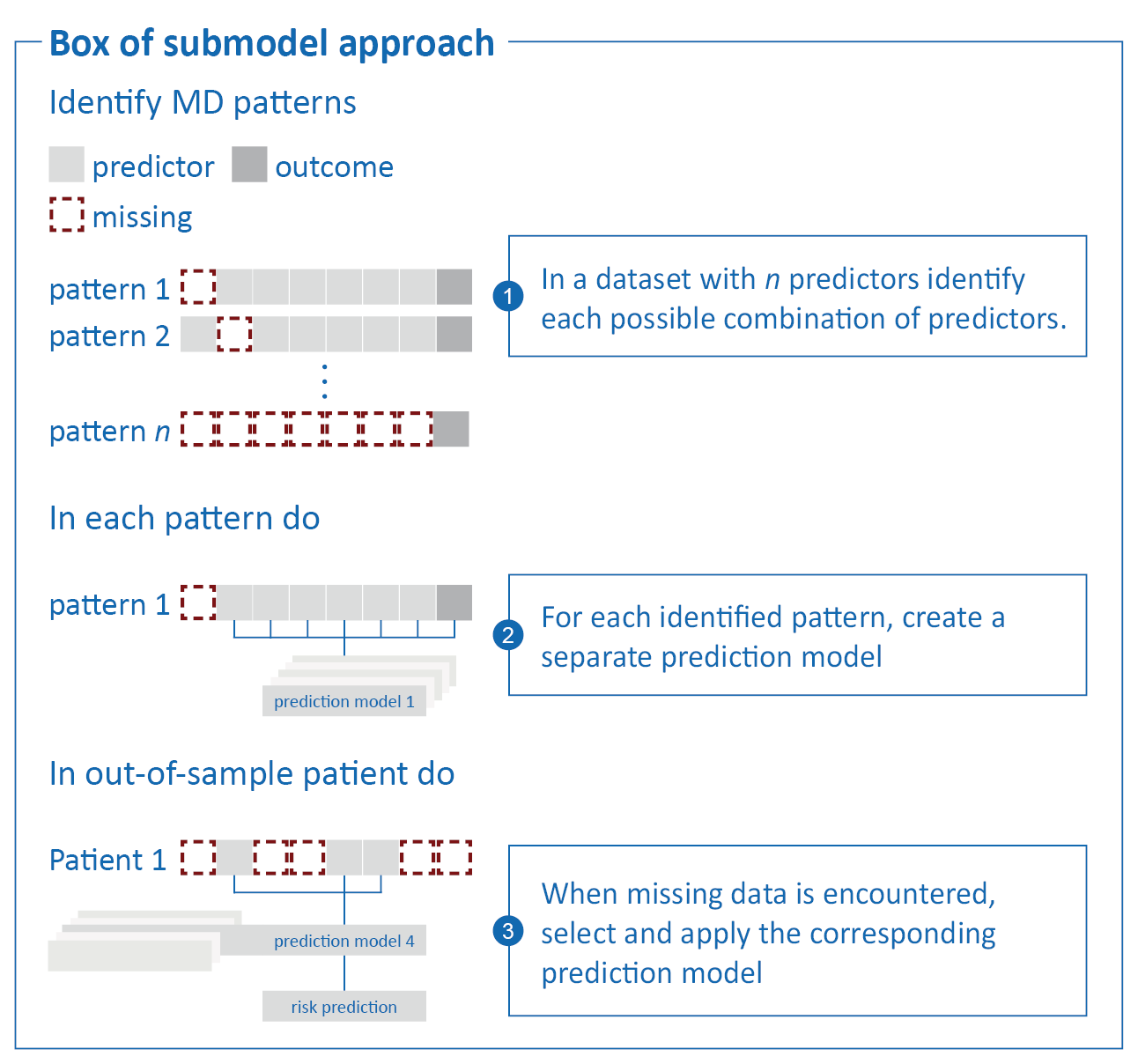
[Need to say something that you develop 2 models: one (traditional) risk prediction model, and one model for multilple imputation]. JMI is an imputation method that involves estimating the multivariate (joint) density of the data and to generate imputed values directly from the conditional distribution (14). Because distribution parameters cannot directly be estimated in incomplete data, JMI typically requires the implementation of a Gibbs sampler. Recently, an extension to JMI was proposed to allow for real-time imputation in individual patients (13,15). With the extension the development of a JMI model consists of two separate steps. In the first step, the means and covariance of all predictor variables are estimated in a development sample. Since JMI assumes that every predictor variable is normally distributed, the population characteristics (i.e., means and covariance) can directly be used to generate, or draw, imputations on an individual level. In clinical practice, when a prediction model now encounters missing values, the developed JMI model can be utilized to generate imputations for each of the missing variables. An advantage of JMI is that it can be applied to a previously developed prediction model. See Figure x for a schematic depiction of JMI. [TODO: add explanation of the three types of JMI that we’ll use here or in simulation design].

Graphical user interface, diagram, timeline

Description automatically generated with medium confidence  
*Figure x. Joint Modeling Imputation (JMI)*

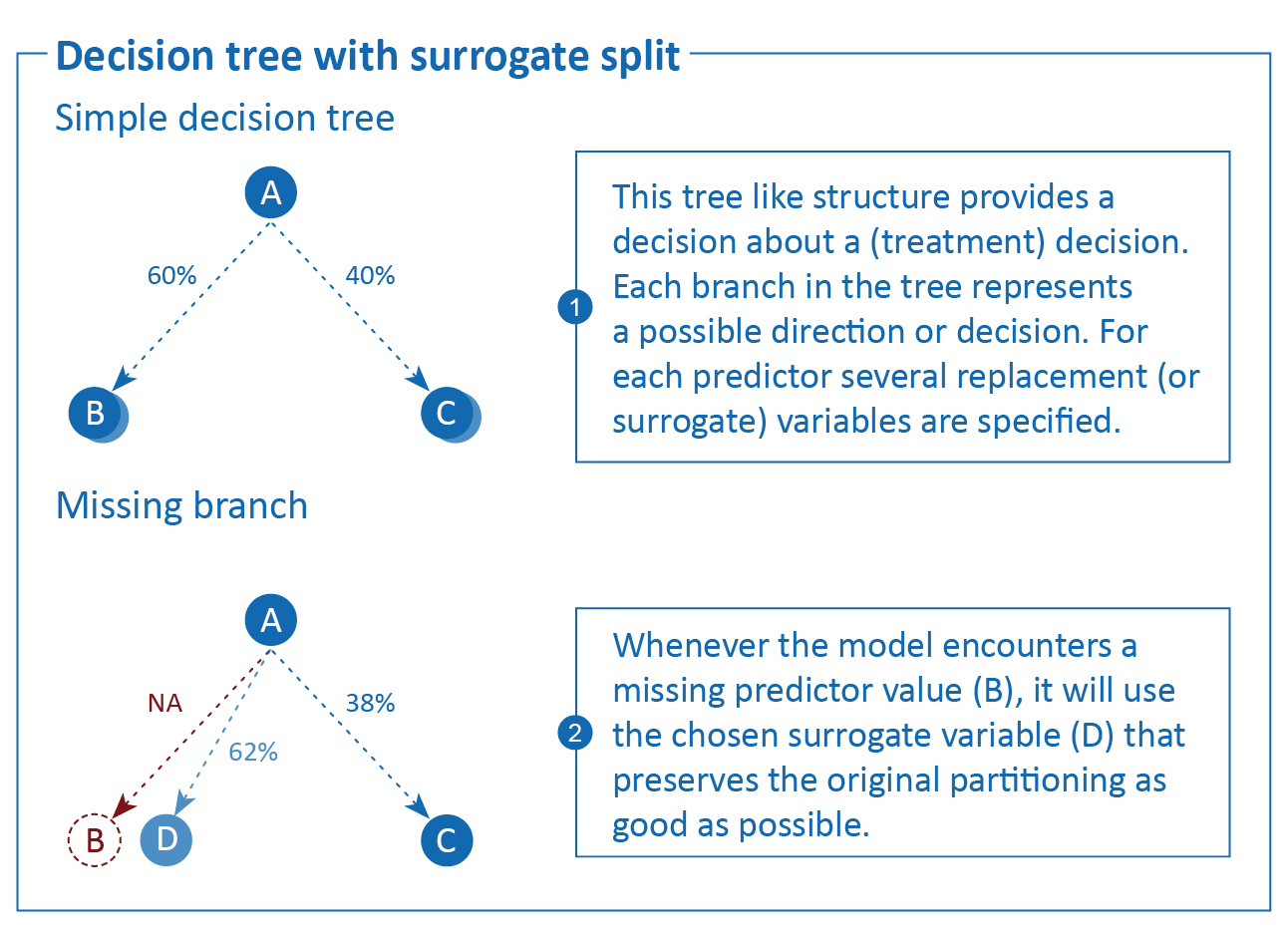
## Pattern Submodel (PS) approach

Another approach to address missing data without requiring imputation is to develop separate prediction models (so called pattern submodels, or briefly, PS) for each missing data pattern (11). Each PS is to be made specifically for one of the identified missing data patterns in the training data and the missing data patterns that are encountered in real-time clinical practice. When applied to a new, out-of-sample, individual, PS approach uses the corresponding prediction model (i.e., matching the missing data pattern at hand). A recent study has shown that the use of pattern submodels for prediction performs similarly to multiple imputation and can be used when the data are missing not at random (MNAR, when missing data is dependent on unobserved values) (11). As such, pattern submodels may provide an elegant and intuitive to understand method for handling missing data when implementing prediction models. See figure x for a schematic depiction of the PS approach.

  
*Figure x. Pattern submodel approach*

## Surrogate Splits (SS)

[first introduce decision trees and random forests briefly, then discuss what iss SS] As an early extension to the well-known decision tree, surrogate splits were developed to circumvent the necessity for imputation (9,10,16). Decision trees use, as the name suggests, a tree like structure to find the optimal cut-off point which partitions the data for optimal predictive performance. Based on the values of the pre-defined predictor variables, each branch in the tree represents a possible direction or decision. In essence, random forests combine multiple decision trees to be merged for improved prediction accuracy. Briefly, surrogate splits try to preserve the partitioning of each original split in a tree as good as possible in the presence of missing predictor values. Whenever the model is applied to an individual and encounters a missing predictor value, it will use the pre-specified surrogate (i.e., replacement) variable, rather than the missing predictor variable, to decide upon the split direction. See figure x for a schematic depiction of surrogate splits in the context of a single decision tree. In this study we use SS in combination with a random forest prediction model (i.e. the aggregate of many decision trees).

*  
Figure x. Decision tree with surrogate splits*

# Simulation design

## Aims

The aim of the simulation study is to emulate how a single patient would present themselves in clinical practice, with incomplete prediction model data, and to evaluate the performance of several on-the-fly missing data handling approaches. We compare the performance of the different missing data approaches w.r.t. their ability to recover missing values and generate accurate risk predictions.. We consider the situation in which a complete dataset is available for prediction model development, and that the resulting model is then applied to individual patients with missing observations for one or more variables. For an overview of the simulation, see Figure x; for the full script and technical details, see github.com/hanneoberman/SIG.

**Diagram

Description automatically generated**

*Figure x. Simulation study*

## Data-generating mechanism

All data are generated from a single model-based population, consisting of ten continuous predictors and one dichotomous outcome. In each simulation iteration, we draw two samples from the population: a complete development set (*n* = 10.000), and a validation set in which we introduce one or more missing values to mimic how patients would present themselves in clinical practice (*n* = 20.000).

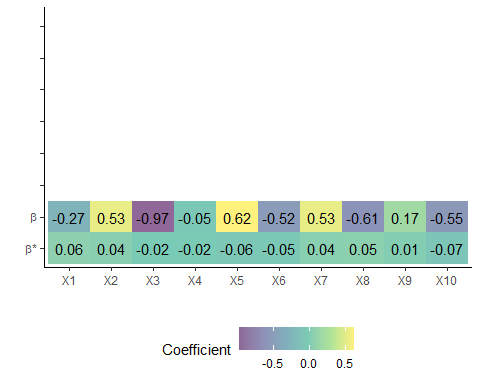
The data generating mechanism of the predictor space is a multivariate normal distribution, , with mean vector and covariance matrix Σ. All 10 predictors have a mean of zero, . The covariance matrix can be found in the Supplementary Materials, and is visualized in Figure XYZ.

|  |
| --- |
| Figure XYZ. Correlation coefficients between predictors |

From the predictor space, we define the binary outcome Y. Y is a function of through the logit link function,

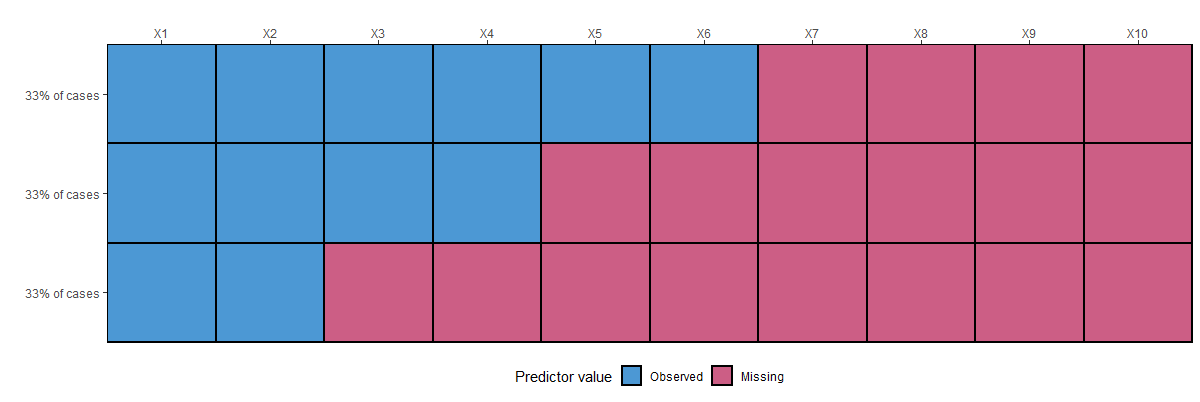
where s are regression coefficients, and residual error . We differentiate between three types of regression coefficients: 1) the intercept, ; 2) a vector of regression coefficients for the main effects of the predictors, ; and 3) an additional vector of regression coefficients for the interactions with the first predictor, . This introduces a polynomial effect of the second degree, , and nine moderation effects. For additional non-linearity, we use a transformation in the effect of the second predictor, . The regression coefficient vectors

are visualized in Figure XYZ.

Figure XYZ. Regression coefficients of the main and interaction effects of the predictors

With an intercept of , the expected outcome occurrence of is 15%.

The validation set is amputed (i.e., made incomplete) according to several missingness mechanisms and missingness rates. In this study, we focus on the Missing At Random (MAR) missingness mechanism (19) [TODO: add MNAR]. We use a mixture of the four kinds of MAR missingness, as described by [REF: Schouten]. The overall missingness rate is 60%, but within each validation set, the missingness rate varies between observations. The hypothetical patients in our validation set are missing either 40%, 60%, or 80% of the observations in the predictor space. The resulting missing data pattern is visualized in Figure x.



**Figure x.** Missing data pattern.

## Estimands

Each row in the validation set represents a hypothetical patient for which we want to predict the absolute risk of the outcome in real-time. Our estimands are the outcome itself (the binary manifestation of Y), and the underlying probability of Y (which is only observable in the context of a simulation study, not in a clinical setting). We estimate Y and the probability of Y from the incomplete predictor space of each validation set.

## Methods

Our methods consist of nine pairs of missing data methods and prediction models. For an overview of all methods, see Table 1.

**Missing data handling strategies.** To accommodate for missing predictor values in real-time, we consider three types of missing data handling strategies: JMI, pattern submodels (PS), and surrogate splits (SS). Since JMI can have different implementations, we further subdivide this strategy into (i) imputing the conditional mean (JMI-CM), (ii) single imputation with a random draw from the conditional multivariate distribution (JMI-SD), and (iii) multiple imputation with 50 draws from the conditional multivariate distribution and pooling (i.e., taking the average of) the predictions of the outcome (JMI-MD).

**Prediction models.** We obtain predictions of the outcome by applying two models on the incomplete (imputed) predictor space. The first prediction model is flexible logistic regression (FLR) with a natural cubic spline. The second prediction model is a random forest (RF). Technical details such as model tuning can be found in the Supplementary Materials and on github.com/hanneoberman/SIG. Both prediction models are compatible with the JMI missing data strategy and pattern submodels missing data strategy. The surrogate split missing data strategy is only available for tree-based prediction models, such as a random forest.

|  |  |  |  |
| --- | --- | --- | --- |
|  | | Flexible logistic regression | Random forest |
| JMI | Conditional mean imputation1 | X | X |
| Single draw imputation2 | X | X |
| Multiple draw imputation3 | X | X |
| Pattern submodels4 | | X | X |
| Surrogate splits5 | |  | X |

**Table 1.** Overview of missing data methods and prediction models  
  
1. Missing values are imputed by the predictor mean, conditional on the observed values of the other predictors; 2. Missing values are imputed by a random draw from the conditional multivariate distribution of the predictor; 3. Missing values are imputed 50 times by a random draw from the multivariate normal distribution, and subsequently used to obtain 50 predictions of the outcome, which are then averaged to obtain one pooled prediction; 4. Missing values are circumvented by selecting the appropriate pattern submodel for predicting the outcome. 5. Missing values are accommodated using surrogate splits.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | Missing data technique | Prediction model | |
| FLR | RF |
| JMI-CM | Conditional mean imputation. Missing values are imputed by the predictor mean, conditional on the observed values of the other predictors | | x | x |
| JMI-SD | Single draw imputation. Missing values are imputed by a random draw from the conditional multivariate distribution of the predictor | | x | x |
| JMI-MD | Multiple draw imputation. Missing values are imputed 50 times by a random draw from the multivariate normal distribution, and subsequently used to obtain 50 predictions of the outcome, which are then averaged to obtain one pooled prediction | | x | x |
| PS | Pattern submodels. Missing values are circumvented by selecting the appropriate pattern submodel for predicting the outcome | | x | x |
| SS | Surrogate splits. Missing values are accommodated using surrogate splits | |  | x |

**Table 1.** Overview of missing data methods and prediction models.

## Performance measures

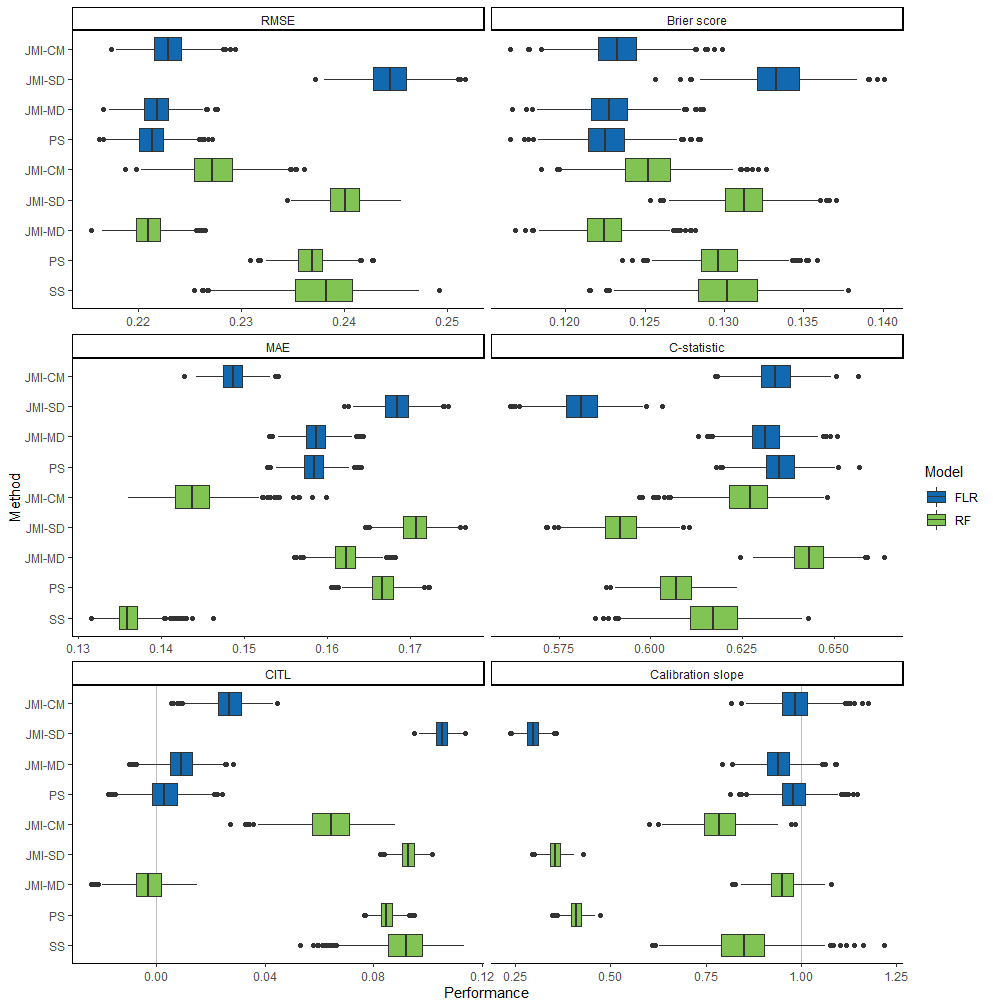
We evaluate the estimates (the predicted risk of the outcome for each of the hypothetical patients) in terms of prediction accuracy at the individual patient-level, and in terms of discrimination and calibration. Subsequently, all metrics are averaged across simulation iterations. Table 2 provides an overview of the performance measures: (i) root mean squared error (RMSE) of the predicted risk, (ii) brier score, (iii) mean absolute error, (iv) concordance (C-) statistic, (v) calibration-in-the-large (CITL) and (vi) the calibration slope.

|  |  |
| --- | --- |
| Measure | Performance metric |
| Prediction accuracy | Root mean square error (RMSE). The RMSE of the predictions reflects the difference between the estimated probability of Y and the true underlying probability of the outcome before amputation. Like the estimand and estimates, the RMSE lies on the probability scale. Lower values indicate better performance (20). |
| Brier score. The brier score is defined as the squared difference between the predicted risk and the true (binary) outcome value. A brier score of 0 would represent a perfect model, whilst the maximum brier score is determined by the incidence of the outcome (20). |
| Mean absolute error (MAE). The MAE is another prediction-level accuracy metric, similar to the RMSE. A lower mean absolute error suggests a better model [REF]. |
| Discrimination | Concordance (C-)statistic. The C-statistic is a rank-order statistic, which is used to describe how well a classification model can discriminate between those with an event and those without. The C-statistic shows the probability of taking two random subjects (one with and one without the outcome) and correctly attributing the one with the outcome with a high risk. A C-statistic of 0.5 describes a model with no discriminative performance and a C-statistic 1 describes a model with perfect discriminative performance. |
| Calibration | Calibration-in-the-large (CITL). The CITL represents the overall calibration of a model. In other words, the extent of agreement between the average predicted risk and the original predicted risk (21). The metric ultimately describes the amount of systematic over- or under-estimation of the predicted risk. A value of 0 is ideal and represents perfect agreement. |
| The calibration slope. In contrast with the CITL, the calibration slope does not evaluate the average predicted, or original, risk. Rather, it quantifies the extent by which the predicted risks vary too much (i.e., slope <1) or too little (i.e., slope >1). Ideally, the slope is 1. |

**Table 2.** Performance measures

Results

Results of the simulation study are visualized in Figure x. Table XYZ presents average performances across simulations. [TODO: explain that JMI-SD is overall the worst and not relevant to interpret?] [TODO: add reference to MNAR performance]



**Figure x.** Performance measures per method  
Note. JMI-CM: conditional mean imputation; JMI-SD: single draw imputation; JMI-MD: multiple draw imputation; PS: pattern submodels; SS: surrogate splits; AUC: area under the curve; MAE: mean absolute error; RMSE: root mean squared error; FLR: flexible logistic regression; RF: random forest

## Root mean squared error

The pattern submodel approach (PS) was best able to recover the original probability of the outcome, especially when implemented with a logistic regression model (rather than a random forest). Similarl performance was obtained when adopting JMI-CM or JMI-MD (again in combination with a LR model). For a random forest prediction model, JMI-MD outperforms all other approaches, including JMI-CM. Especially surrogate splits and PS show relatively low accuracy (in addition to the overall worst performing technique JMI-SD).

## Brier score

PS, similar to the RMSE, can best approximate the binary realization of the outcome when paired with a flexible logistic regression model. Again, JMI-CM and JMI-MD closely match the performance of PS. When a random forest prediction model is used, JMI-MD is superior to all other approaches. Once more, surrogate splits and PS have relatively poor performance.

## MAE

[TODO: discuss whether it is informative to interpret this metric if it doesn’t suit our DGM]. Yes, please include. Interpretation of MAE may warrant some discussion

## C-index

Again, when utilized for a flexible logistic regression model, PS exceed the performance of other techniques, now in terms of discriminating between cases and non-cases. The discriminatory ability of JMI-CM and JMI-MD are equivalent to PS. The performances of JMI-CM and PS are, once more, diminished when a random forest prediction model is used instead of FLR. And, although slightly better than PS, the performance of surrogate splits is below par.

## Calibration-in-the-large

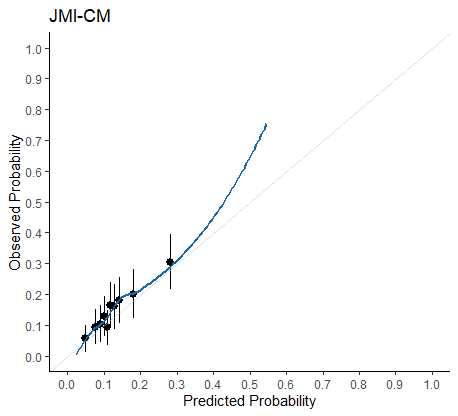
In terms of CITL, PS only slightly outperforms JMI-CM and JMI-MD with all showing near perfect overall calibration when paired with a flexible logistic regression model. All missing data handling techniques show similar calibration when paired with a random forest prediction model. The clear favourite is JMI-MD with near perfect CITL.

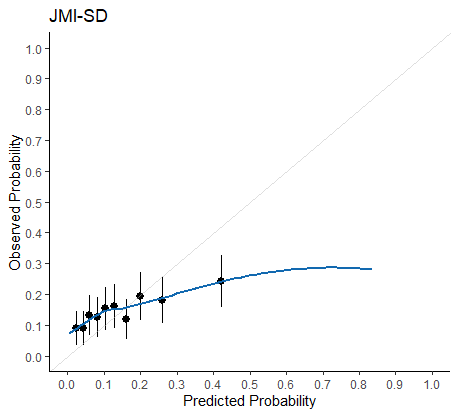
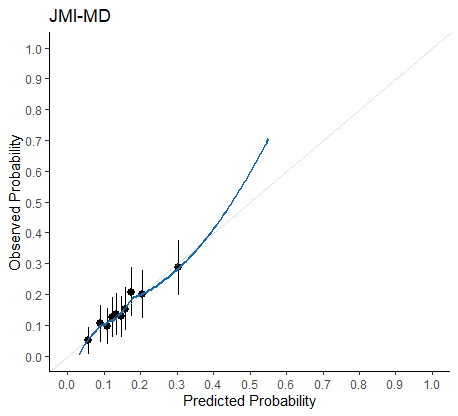
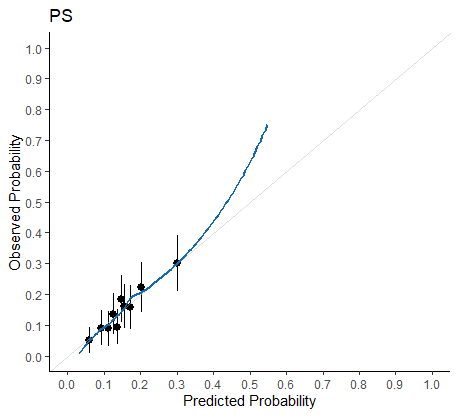
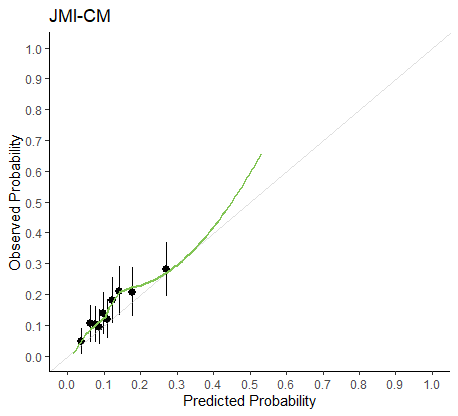
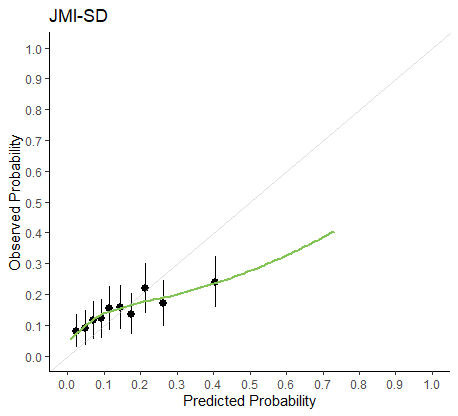
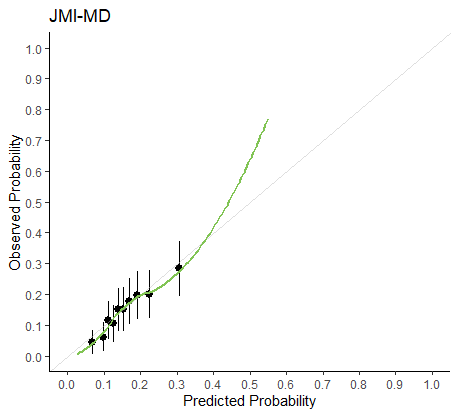
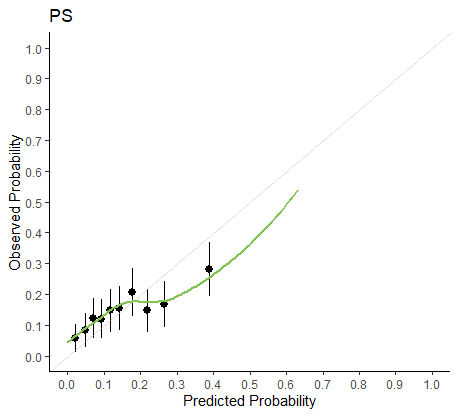
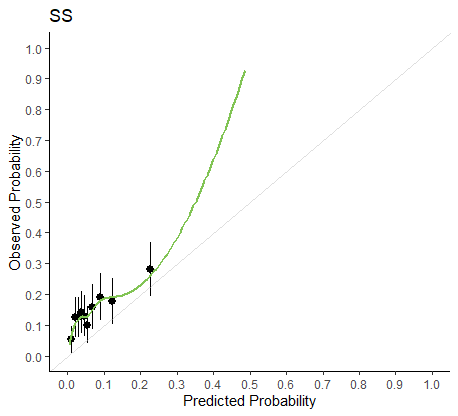
## Calibration slope

In contrast with other performance metrics, the use of JMI-CM in a flexible logistic regression can best quantify the extremeness of predicted risks across the whole range. Both JMI-MD and PS have very similar performance. Both With the exception of JMI-MD, all missing data handling techniques show miscalibration when a random forest prediction model is used.

## Calibration plots [TODO: interpret figures or move to supplement]

The next figures show the calibration plots for a subset of 1000 patients. Grey lines are optimal calibration, cubes show average performance for the deciles, with 95% CI as error bar. Colored lines (blue for FLR, green for RF) are Loess lines through the calibration.



# Discussion

This simulation study aimed to evaluate the effectivity of using real-time missing data handling strategies to handle missing predictor values in individual patients. We considered JMI, submodels and surrogate splits for the real-time handling of missing data when using either a flexible logistic regression or random forest model. Our simulation study showed that the optimal choice of missing data handling technique may be dependent on the preferred prediction modelling approach. The pattern submodels approach does not involve any imputation and generally leads to adequate prediction model performance in the presence of missing data. A major advantage is that PS methods do not require MAR assumptions, and therefore offer an appealing solution in real-world datasets. However, when PS is implemented with RF, problems arise with [….]. Possibly, RF are more prone to overfitting when estimated in smaller (sub)samples as compared to logistic regression. [Maybe introduce here the only other non-imputation approach: SS. Although it is conceptually a bit similar to PS (and more tailored to trees than PS), it does not give very good performance. Discuss why this could be the case]

Finally, when adopting imputation methods to accompany the (RF or logistic regression ) model , simulation results indicate that JM works reasonably well, provided that multiple imputations are generated for each missing value.

For random forest prediction models, JMI showed the best performance depending on the implementation. For example, multiple imputation performed more consistently than imputing the conditional mean and single draws severely underperformed on all metrics. Our results suggest that built-in mechanisms such as surrogate splits perform relatively poorly, when compared with the other missing data handling approaches.

An important limitation to our simulation study is the choice of DGM. Based on previous studies, moderate correlations between predictor variables were imposed on the predictor space [REF: NIjman]. This limits the potential usefulness of imputation methods, as their implementation relies on the presence of correlations in the observed data. Previously, low correlations have been associated with limited performance of JMI (12). Similarly, surrogate splits are highly dependent on the correlation between the missing predictor value and the surrogate replacement value (22). Hence, the choice of DGM may explain the deficient performance of surrogate splits.

To avoid overfitting, prediction models are typically designed as simple as possible, and therefore only include predictors that are not strongly correlated to one another. . It is therefore important to consider imputation strategies that have access to all relevant variables, even if they are not used for generating risk predictions. In our simulation study, we only generated 10 covariates, all of which were used for development of the prediction model and imputation strategies. In practice, however, many more variables may be available. It is therefore possible that some variables may only be helpful for generating imputations, but not for risk prediction. Deciding which variables to include in an imputation model (such as JMI) is, however, not straightforward. [Previosu studies showed that …]

The submodel approach is not susceptible to performance loss when predictor variables have low correlations. Although this seems to be an advantage over the other methods evaluated in this simulation, the performance become much worse when a random forest prediction model is used. This may be explained by the fact that less predictors ultimately restrict how much a random forest may vary between each tree (23). In other words, if there are less features available, as is the case in a submodel approach, the number of possible trees is limited.

Finally, we found that logistic regression generally yielded better performance than RF. Possibly, this is because our dataset included mostly continuous predictors, and did not have very high dimension. RF are known to perform particularly well when dealing with a very large amount of discrete variables (especially in the presence of interactions), since it naturally allows for partitioning and variable selection. However, within the RF approach, we found that multiple imputation with JMI yielded the best performance.

TODO: RF en GLM uit elkaar houden. Deze data is dus blijkbaar niet ideaal voor ML. Extra limitatie is dat de twee non-parametrische variabelen (X1 en X2) zijn altijd geobserveerd.

[TODO: laatste alinea met pakkende uitsmijter toevoegen:

* de beste missing data method hangt af van welk prediction model gebruikt wordt.
* een veilige keuze is multiple draw JMI.
* single draw imputation NIET gebruiken]

In short, the best missing data handling techniques for the flexible logistic regression prediction model are pattern submodels, multiple draw imputation and conditional mean imputation; the best technique for the random forest model is multiple draw imputation.

Multiple draw imputation has the best performance across the two prediction models, but pattern submodels and conditional mean imputation work well with flexible logistic regression models as well.

JMI-SD is disqualified.

# References

1. Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM. Review: A gentle introduction to imputation of missing values. Journal of Clinical Epidemiology. 2006 Oct;59(10):1087–91.

2. Janssen KJM, Vergouwe Y, Donders ART, Harrell FE, Chen Q, Grobbee DE, et al. Dealing with Missing Predictor Values When Applying Clinical Prediction Models. Clinical Chemistry. 2009 May 1;55(5):994–1001.

3. D’Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General Cardiovascular Risk Profile for Use in Primary Care: The Framingham Heart Study. Circulation. 2008 Feb 12;117(6):743–53.

4. Dorresteijn JAN, Visseren FLJ, Wassink AMJ, Gondrie MJA, Steyerberg EW, Ridker PM, et al. Development and validation of a prediction rule for recurrent vascular events based on a cohort study of patients with arterial disease: the SMART risk score. Heart. 2013 Jun 15;99(12):866–72.

5. Piepoli MF, Hoes AW, Agewall S, Albus C, Brotons C, Catapano AL, et al. 2016 European Guidelines on cardiovascular disease prevention in clinical practice: The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts)Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). Eur Heart J. 2016 Aug 1;37(29):2315–81.

6. Groenhof TKJ, Bots ML, Brandjes M, Jacobs JJL, Grobbee DE, van Solinge WW, et al. A computerised decision support system for cardiovascular risk management ‘live’ in the electronic health record environment: development, validation and implementation—the Utrecht Cardiovascular Cohort Initiative. Neth Heart J. 2019 Sep;27(9):435–42.

7. Bezemer T, de Groot MC, Blasse E, ten Berg MJ, Kappen TH, Bredenoord AL, et al. A Human(e) Factor in Clinical Decision Support Systems. J Med Internet Res. 2019 Mar 19;21(3):e11732.

8. Van Buuren S. Flexible Imputation of Missing Data. 2nd ed. CRC Press; 2018.

9. Feelders A. Handling Missing Data in Trees: Surrogate Splits or Statistical Imputation? In: Żytkow JM, Rauch J, editors. Principles of Data Mining and Knowledge Discovery [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 1999 [cited 2019 Oct 2]. p. 329–34. Available from: http://link.springer.com/10.1007/978-3-540-48247-5\_38

10. Cevallos Valdiviezo H, Van Aelst S. Tree-based prediction on incomplete data using imputation or surrogate decisions. Information Sciences. 2015 Aug;311:163–81.

11. Fletcher Mercaldo S, Blume JD. Missing data and prediction: the pattern submodel. Biostatistics. 2020 Apr 1;21(2):236–52.

12. Nijman SWJ, Hoogland J, Groenhof TKJ, Brandjes M, Jacobs JJL, Bots ML, et al. Real-time imputation of missing predictor values in clinical practice. European Heart Journal - Digital Health. 2021 May 4;2(1):154–64.

13. Nijman SWJ, Groenhof TKJ, Hoogland J, Bots ML, Brandjes M, Jacobs JJL, et al. Real-time imputation of missing predictor values improved the application of prediction models in daily practice. Journal of Clinical Epidemiology. 2021 Jun;134:22–34.

14. Nijman S, Groenhof T, Hoogland J, Bots M, Brandjes M, Jacobs J, et al. Real-time handling of missing predictor values when implementing and using prediction models in daily practice. JCE. 2021;Article in press.

15. Hoogland J, Barreveld M, Debray TPA, Reitsma JB, Verstraelen TE, Dijkgraaf MGW, et al. Handling missing predictor values when validating and applying a prediction model to new patients. Statistics in Medicine. 2020 Jul 20;sim.8682.

16. Hapfelmeier A. Analysis of Missing Data with Random Forests [Internet]. 2012 [cited 2019 Sep 4]. 6–7 p. Available from: https://edoc.ub.uni-muenchen.de/15058/1/Hapfelmeier\_Alexander.pdf

17. Johnson AEW, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data. 2016 Dec 20;3(1):160035.

18. Jones AE, Trzeciak S, Kline JA. The Sequential Organ Failure Assessment score for predicting outcome in patients with severe sepsis and evidence of hypoperfusion at the time of emergency department presentation\*: Critical Care Medicine. 2009 May;37(5):1649–54.

19. Rubin DB. Inference and Missing Data. Biometrika. 1976;63(3):581–92.

20. Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. Berlin: Springer; 2009. 497 p. (Statistics for biology and health).

21. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. European Heart Journal. 2014 Aug 1;35(29):1925–31.

22. Twala B. AN EMPIRICAL COMPARISON OF TECHNIQUES FOR HANDLING INCOMPLETE DATA USING DECISION TREES. Applied Artificial Intelligence. 2009 May 4;23(5):373–405.

23. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak. 2019 Dec;19(1):281.

**Supplementary Materials (just here to keep the formatting, real one will be a separate file)**