Missing Data in Prediction Research

JOURNAL OF NEUROTRAUMA 38:1842–1857 (July 1, 2021)

DOI: 10.1089/neu.2020.7218

Missing Data in Prediction Research: A Five-Step Approach for Multiple Imputation, Illustrated in the CENTER-TBI Study

Benjamin Yaël Gravesteijn,[1] Charlie Aletta Sewalt,[1] Esmee Venema,[1] Daan Nieboer,[1] and Ewout W. Steyerberg[1,2]; and the CENTER-TBI Collaborators*

Abstract

In medical research, missing data is common. In acute diseases, such as traumatic brain injury (TBI), even well-conducted prospective studies may suffer from missing data in baseline characteristics and outcomes. Statistical models may simply drop patients with any missing values, potentially leaving a selected subset of the original cohort. Imputation is widely accepted by methodologists as an appropriate way to deal with missing data. We aim to provide practical guidance on handling missing data for prediction modeling. We hereto propose a five-step approach, centered around single and multiple imputation: 1) explore the missing data patterns; 2) choose a method of imputation; 3) perform imputation; 4) assess diagnostics of the imputation; and 5) analyze the imputed data sets. We illustrate these five steps with the estimation and validation of the IMPACT (International Mission on Prognosis and Analysis of Clinical Trials in Traumatic Brain Injury) prognostic model in 1375 patients from the CENTER-TBI database, included in 53 centers across 17 countries, with moderate or severe TBI in the prospective European CENTER-TBI study. Future prediction modeling studies in acute diseases may benefit from following the suggested five steps for optimal statistical analysis and interpretation, after maximal effort has been made to minimize missing data.

Keywords: imputation; missing data; prediction; traumatic brain injury; tutorial

Introduction

Missing data is a common problem in medical research.[1] Missing data occurs in studies with routinely collected data, and even if data are prospectively collected with attempts to minimize the occurrence of missing data (Box 1).

Box 1: Exemplary Causes of Missing Data

Missing data occurs for example when there is no response to surveys or follow-up appointments; or if laboratory tests or imaging were not ordered for all patients; or when a score or test was simplified instead of properly executed to save time at a busy emergency department. For example, the Glasgow Coma Scale (GCS) may be noted as 13 instead of the detailed score (e.g., E5M6V2); or ''Pupils Equal And Reactive to Light'' instead of the exact diameters of the pupils with and without exposure to light.

Prediction modeling is central to many domains of medicine, such as screening, diagnostics, and therapy. It is a growing research area.[2,3] Predictions are based on the combination of characteristics for a diagnostic outcome (e.g., presence of abnormalities at computed tomography [CT] scan) or prognostic outcome (e.g., Glasgow Outcome Scale [GOS] at 6 months).[3] Potential predictors commonly relate to the characteristics of the patient, disease, or treatment.[4] Statistical models typically drop patients with any missing values from analyses. Prediction research is particularly sensitive to the occurrence of missing data, because it relies on the statistical combination of multiple variables, which may each have missing values.

Analyzing only the available data (often referred to as a ''complete case analysis'') is the most basic approach to deal with missing data. It decreases the available information for statistical analyses. Moreover, it could potentially introduce selection bias given that patients with observed characteristics may be systematically different from patients with missing values.[2,5,6] For example, patients with missing baseline characteristics may be a selected subgroup, because laboratory tests or imaging might not be ordered for less-severe cases.

Although the problem of missing data is complex, some methodological standards to deal with missing data are available. General recommendations to handle missing data have been suggested.[7] One generic recommendation is that multiple imputation is the method of choice in many areas of medical research.[8] Imputation exploits the availability of information from non-missing predictors for partly complete patients rather than discarding these

[1]Department of Public Health, Erasmus Medical Center, Rotterdam, The Netherlands.
[2]Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands.
*The CENTER-TBI Collaborators may be found at the end of this article.

patients. A recent systematic review showed that missing data is reported inconsistently and often handled suboptimally, underscoring the importance of a practical framework.[9]

We aim to provide guidance to deal with missing data in prediction research. Some nuances and methodological considerations are provided in text boxes. We first address various general issues in prediction modeling and handling of missing data, specifically with imputation procedures. We then propose a five-step approach to perform imputation as a way to deal with missing data, illustrated with a case study in a multi-national, prospective cohort study for traumatic brain injury (TBI): CENTER-TBI.[10,11]

Principles of prediction modeling

Prediction modeling entails the prediction of a clinically relevant outcome based on the combination of multiple predictors' effects.[12] Prediction models typically provide estimates of the relative effect of predictors in the model and absolute risk predictions for individual patients. The CRASH (Corticosteroid Randomisation After Significant Head Injury) and IMPACT (International Mission on Prognosis and Analysis of Clinical Trials in Traumatic Brain Injury) models are examples of such prediction models in TBI.[13,14] These models predict mortality and 6 months' unfavorable outcome according to the GOS.[13,15,16]

The performance of such models is ideally assessed in external validation studies, where predictions of an existing model are compared to observed outcomes in a new setting. For binary outcomes, the performance is commonly assessed using discrimination and calibration.[8] Discrimination refers to the ability of a prediction model to discriminate between patients with and without the outcome of interest. It is commonly assessed by the area under the receiver operating characteristic curve (AUC; or c-statistic). Calibration refers to the agreement of predicted probabilities of a model and observed outcomes (e.g., ''if the risk of death is x%, do x% of the patients with this prediction actually die?''). Calibration can be assessed graphically and can be summarized according to calibration-in-the-large (measuring under- or overestimation of the average predicted risk) and calibration slope (measuring the average strength of overall predictor effects).[17]

The current tutorial gives guidance on how to deal with missing data to use all available information in a data set to develop a prediction model and validate an existing model. Effectively, this boils down to correct estimation of the parameters in the model at model development. At validation, we apply the model to new data and compare the observed outcome to the predicted risk.

Mechanisms of missing data

To describe missing data, a paradigm of three distinctive mechanisms for missing data has been established (Table 1).[5] First, Missing Completely At Random (MCAR) arises when missingness is not associated with observed or unobserved variables. For example, this missingness arises when administrative errors or accidents occur. Second, Missing At Random (MAR) is defined as missingness that is associated with observed variables. For example, patients with lower injury severity scores may have more missing CT scans. Finally, Missing Not At Random (MNAR) arises when the missingness is associated with unobserved variables or the value of the variable itself. For example, patients may be less likely to fill in their income in a survey if their income is substantially higher than average. Both missingness in the predictors and the outcome can be categorized according to these missing data mechanisms. An example for MNAR in the outcome is that patients may not return for follow-up visits if they are doing very poorly or very well.

Methods for imputation

The most basic method for dealing with missing data is a complete case analysis. This leads to loss of statistical power[18] and, potentially, to bias in risk predictions.[18] The most problematic consequence of less statistical power is a higher risk of overfitting of a prediction model. The model performs much better in the original data set (too ''optimistic'') than the model will perform when it is used to calculate risk for new patients. Bias in risk prediction as a result of a complete case analysis may also result in a systematically over- or underestimated risk (poor calibration-in-the-large). Therefore, many methodologists and journal editors currently recommend to statistically impute missing data.[8,19] For such imputation, we have to assume that the mechanism of missingness is MCAR or MAR, not MNAR. Further, we assume that the imputation model, which only includes observed variables, is valid. We will discuss variants of single imputation and multiple imputation (Table 2). Multiple imputation, because of its theoretical attractiveness, has become a methodological standard. But as every method has its limitations,[20] an educated and balanced choice based on the research aim should be made.

Single imputation. Replacing the missing value by a different value is referred to as single imputation. We describe two distinct methods for replacing this value: average imputation and single imputation using conditional estimates. The theoretical disadvantage of single imputation in general is that the uncertainty in imputation is not incorporated in the final analysis (Box 2).

Imputing the value by the average value as the best estimate for the missing value may translate to taking the mean for normally distributed variables, median for non-normally distributed variables, and mode for categorical variables. This method typically results in biased estimates of parameters under any missing data mechanism at model development (Box 3).

Table 1. Missing Data Mechanisms for Predictors with Examples[2]

Label	*Missing mechanism*	*Description*	*Clinical example*
MCAR	Missing completely at random	Administrative errors, accidents	A batch of blood samples got lost.
MAR	Missing at random	Missingness related to known patient characteristics	Older patients have more difficulty to come to follow-up visits.
MNAR	Missing not at random	Missingness related to 1) the value of the predictor, or 2) to variables not available in the analysis.	1) Self-report of income in a survey: lower incomes are less likely to be reported. 2) Patients with impaired cognition (not measured) cannot understand to fill in a neuropsychological test.

Table 2. Methods of Dealing with Missing Data

Method	*Description*	*Valid[a] under*
Complete case	Drop cases with missing values.	MCAR
Average imputation	Replace the missing values by the average.	—
Single imputation by conditional estimation	Replace the missing values by the most likely value based on the observed data.	MCAR, MAR
Multiple imputation	Replace the missing values by the most likely values based on the observed data, multiple times.	MCAR, MAR
Multiple imputation, including the outcome	Multiple imputation, also imputing the outcome	MCAR, MAR

[a]valid here is defined as providing unbiased estimates of the final parameters in the model.

Single imputation by conditional estimates translates into using predictions of the most likely value of the missing observation from a regression model. Both categorical and continuous predictors may be estimated by a variety of regression models (see step 3: performing imputation). Such conditional imputation provides better point estimates of the prediction model than average imputation. We may also draw imputed values from a distribution of likely values.[1] Different imputed data sets then yield different estimates in the statistical analysis. Although the random drawing of values now incorporates some uncertainty, the final analysis still treats these imputed values as observed. Because this technique incorporates existing correlations in observed data, it is not only valid under MCAR, but also under MAR.

Box 2: Single Imputation and Uncertainty

The main theoretical disadvantage of single imputation is that the uncertainty of the imputed values is not fully taken into account in the estimation of the final model at development: The imputed values are used as observed values in the final analysis. The result is that the standard errors (a measure of uncertainty) are too low. However, it can be argued that this disadvantage is not relevant. If our main aim is to predict the most likely risk for a new patient, the reliability of that prediction is more important than the uncertainty in the estimated parameters. Notably, machine learning algorithms do not incorporate uncertainty into their parameter estimates and are only compatible with single data sets. Single imputation may also be considered when working with very large sample sizes to reduce the computational time.

Box 3: Average Imputation and Bias toward the Null

Average imputation seems like a non-educated guess for the missing observation: The distribution can create a non-natural spike of the mean. This simple imputation approach leads to bias in the estimated predictor effects. This bias is usually toward the null, which makes the approach generally conservative. This behavior is similar to shrinkage of regression coefficients toward the null, as can be achieved with penalized regression methods.[21] Average imputation biases the parameters toward the null even when the sample size is large. Therefore, penalized regression models are preferred to address overfitting rather than average imputation.

Multiple imputation. Multiple imputation is an extension of single imputation using conditional estimates and is valid under the same missing data mechanisms (MCAR, MAR).[1] Instead of imputing one value per patient per missing value, multiple values are drawn with models fitted on the observed variables and stored in multiple completed data sets. The statistical analysis is then performed on each imputed data set separately, and the results can be pooled using specific rules (''Rubin's rules''; Box 4).[22]

Box 4: Rubin's Rules for Prediction Research

Estimates of regression coefficients and performance measures are simply averaged. For the standard error, the between-imputation variance of the coefficients and performance measures is added to the average variance within imputation sets. Therefore, multiple imputation leads to a better estimation of variability in the parameters than single imputation.[5,6,22] Modern software makes it relatively easy to implement multiple imputation for prediction models based on regression techniques.[22] The situation is more complex for machine learning algorithms, where pooling across data sets is usually not so easy. For example, classification trees will differ between imputed data sets and cannot be combined directly.

Methods: Case Study

As a case study, we consider updating (or re-estimation) and external validation of the IMPACT prognostic model in the CENTER-TBI study. The IMPACT model aims to predict 6-month mortality and unfavorable outcome in moderate and severe TBI patients.[13] With good discriminatory performance (AUC $\sim$0.80[13]) and multiple external validation studies,[23–26] the IMPACT model might be considered a rather robust prediction model. The full model contains 10 predictors, with 18 logistic regression coefficients (Supplementary Appendix SA, Table SA1).[27] It was developed in the IMPACT data base, a collection of eight trials and three surveys, with a total of 8509 patients.[28]

We validated this model in the CENTER-TBI data base, a European multi-national, prospective cohort study including all severities of TBI.[10,11] We selected 1375 patients with moderate (GCS 9–12) and severe (GCS 3–8) TBI (Supplementary Appendix SA, Table SA1), included in 53 centers across 17 countries.

The analyses described in this article were all performed in the freely available R software (A language and environment for statistical computing; R Foundation for Statistical Computing, Vienna, Austria). The code with explanations is available in Supplementary Appendix SC.

Results: Five Steps for Missing Values

We consider five steps to deal with missing data in prediction research (Table 3).

Table 3. A Five-Step Approach to Deal with Missing Data in Prediction Research

Step	*Action*	*Main objective*
1	Explore missingness	Assess the quantity and pattern of missingness.
2	Choose imputation method	Balance the benefits and harms of alternative imputation methods in relation to the prediction question.
3	Perform imputation	To produce one or multiple imputed data sets
4	Diagnostics	Assess the convergence of the iteration process, and compare the imputed data with the original data set.
5	Analyze	Estimate the regression coefficients in a prediction model, or assess performance for an existing prediction model.

Step 1: Explore missingness

The first step in handling missing data is inspecting the data set and exploring missingness of variables (Table 3). Recommended steps are to assess:

1. The quantity of missingness: the proportion of missing observations in each variable
2. Patterns of missingness: the frequency of specific missing data patterns per patient
3. Correlation between variables
4. Associations between missingness in one variable and other variables

One of the main considerations for missing data is what proportion of missingness in a predictor is still acceptable when imputing missing data. The stability of imputation declines with higher proportions of missingness.[29] A general advice is that the researcher should decide what proportion can be accepted based on their research question and their context (Box 5).

Box 5: Maximum Proportion of Missingness to Impute

No consensus has been reached about a maximum limit of missingness per variable or per patient. Limits may depend on the specific research question and context.[2] For example, when we are interested in the diagnostic value of a specific biomarker, we would probably not impute missing values for the biomarker and be liberal in imputing missing values for other covariates that may potentially act as confounders. In contrast, when we consider prediction based on the combination of predictors, we may focus on including strong predictors with few missings (which are imputed). Depending on what assumption the researcher is willing to make, multiple imputation can be used for different situations: If MCAR or MAR is plausible, multiply imputing $\sim$10–20% missing values per variable is generally acceptable, whereas under MCAR perhaps 50% missing values can be imputed without much instability in predicted values.[2] The larger the proportion of missing values, the more important that you specified the imputation model correctly, or else the final parameters might be biased.[29] Nevertheless, when the model is correctly specified, multiple imputation still reduces bias compared to a complete case analysis.[30] Multiple imputation accounts for the uncertainty of the imputed values in the standard errors of the final model, but relies on the validity of the imputation model. Higher proportions of missingness motivate a higher number of imputation rounds.[1]

The quantity and pattern of missingness can be visualized (Fig. 1). If a particular pattern per patient arises more frequently than other patterns, a reason should be hypothesized. In our example, glucose and hemoglobin are often missing together in our case study. This might be MCAR: If no blood sample was taken (perhaps by mistake), both glucose and hemoglobin are missing. Similarly, imaging characteristics are also missing together frequently: If the patient did not undergo CT scanning, all these variables are missing together. This could be considered as MAR: Less-severe patients (higher GCS scores) are less likely to undergo a CT scan. Missing data mechanisms should always be assessed critically. Evaluating specific combinations of missingness may help to do so.

Imputation is more efficient if variables are correlated. A correlation plot can be useful (Fig. 2). When highly correlated variables are identified (r close to 1, e.g., hemoglobin and hematocrit), only one of the highly collinear variables is sufficient to use, preferably the variable with the lowest proportion of missingness.

Associations between missingness in one variable and other variables can be investigated to test the MCAR assumption. This can be done by logistic regression analysis with missingness of a variable as the dependent variable. All other variables are the predictors. For example, we observe that the variable ''pupils'' is more often missing in patients with higher motor GCS (Supplementary Appendix SA, Fig. SA1). Although MCAR versus non-MCAR can be tested, the practical value remains unclear: There is no consequence for the consecutive strategy. What would be useful instead is testing MAR versus MNAR. However, this is impossible, because the information to test MAR versus MNAR is missing. Therefore, we still need to make the assumption that the data are MAR, not MNAR, when performing imputation.

To increase the likelihood for the MAR assumption to hold, auxiliary variables may be added to the imputation model. Auxiliary variables should meet two criteria. First, they should be frequently observed: If the auxiliary variable is often missing together with the variables of interest, there is no information gained by adding this variable. Second, the auxiliary variable should contain statistical information about the variables of interest: Variables that are not associated with the predictor variables do not add information to the imputation model. An example of a potential auxiliary in our data set is Injury Severity Score (ISS): It was present for 1365 of 1375 patients (99.3%) in our data set and was correlated with some of the IMPACT predictors (Fig. 2). Moreover, the analysis of variance test of ISS across motor GCS or pupil categories was statistically significant (both, $p < 0.001$). Additionally, adding variables which correlate with other variables might increase the adequacy of the imputation model.[31,32]

In conclusion, exploring missing data should at least assess the quantity and patterns of missingness. This exploration should assess all variables that will be added to the imputation model: all

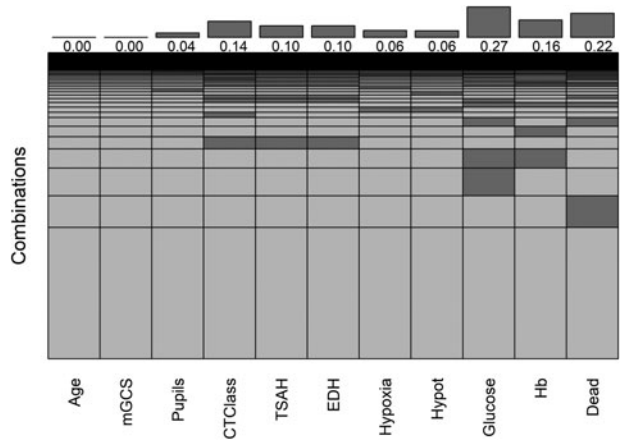

FIG. 1. Quantity and pattern of missingness for each variable in the Center-TBI data base ($n = 1375$). On the top of the graph, the proportion of missingness is shown. In the table, the patterns of missingness are shown, with cell size proportional to number of patients. Each row represents a patient group, and the dark blocks indicate combinations of variables that are missing. A total of 589 of 1375 (43%) had fully complete data. CTClass, Marshall CT class; EDH, epidural hematoma; Hb, hemoglobin; mGCS, modified Glasgow Coma Scale; TSAH, traumatic subarachnoid hemorrhage.

predictors and the outcome. Additionally, we should identify highly correlated variables and select the one that is most relevant and most frequently observed to add to the imputation and prediction model. Moreover, we can test MCAR versus non-MCAR. For imputation of predictors, the result of this test may not impact the strategy to deal with missing values, because we still need to assume that the data are MAR. Imputing the outcome might be especially beneficial when the outcome is non-MCAR (Box 6). Finally, we should focus on adding auxiliary variables in the imputation model, such as ISS in this case study. Adding these variables increases the likelihood of the MAR assumption to hold.

Step 2: Choose method of imputation

The approach to deal with missing data may depend on the size of the data set and the proportion of missing values (Table 4).

With smaller data sets (say total $n < 1000$, or <100 events), multiple imputation is the default method of choice to estimate the relative effects of predictors, with the appropriate associated standard error. However, we might also consider single imputation for prediction of absolute risk, especially if the sample size is large (Box 1).

Variability between imputations may be substantial, especially with a large proportion of missing values. Some researchers might find average imputation attractive, which may be defendable in situations with a very small proportion of missingness.

Step 3: Performing imputation

Single and multiple imputation. For single and multiple imputation, flexible imputation using chained equations is the current standard, for example with the *MICE* package in R[1] or the *mi impute chained* command in Stata. The procedure uses an iterative imputation method, to arrive at stable estimates of imputations for unobserved variables. There are various models within the MICE framework to impute with, the most standard options are Bayesian linear regression (''norm'') for normal distributed variables, predictive mean matching (''pmm'' or ''midastouch'') for non-normally distributed continuous variables, logistic regression (''logreg'') for binary categorical, polytomous regression (''polyreg'') for categorical variables with more than 2 categories, and proportional odds logistic regression (''polr'') for when a categorical variable is ordered.

We draw one imputed value for a single completed data set (single imputation), or create multiple copies of the data set (multiple imputation). Although Rubin famously claimed that three copies were enough for stability,[22] five data sets are now seen as the minimum. Creating even more imputed data sets often might be more beneficial, for example when instability is expected because of substantial proportions of missing values,[1] for example 20 imputed sets if 20% of the values of a predictor are missing.

Finally, it is important to include the outcome in the imputation model. Not including the outcome in the imputation model can result in biased coefficients of the prediction model (Box 6).[33]

Box 6: Imputation for Predictors and/or Outcome

A distinction should be made for imputation of missing values of predictors versus outcomes. It is generally accepted that it is important to include the outcome when imputing predictor values. More controversial is the possibility to impute the outcome. A recent article discusses handling missing outcome data in TBI studies in general.[9] From a predictive modeling perspective, imputation seems counterintuitive: We aim to predict the outcome; how could imputation help in this case?

First and foremost, the average risk estimate might be biased if we exclude patients with missing outcomes: If those patients had on expectation worse outcomes, the prediction model will underestimate the average risk. An exemplary study is the external validation of various CT-decision rules by Foks and colleagues,[34] where the outcome (intracranial abnormality) was imputed to arrive at unbiased estimates of sensitivity and specificity. Excluding patients without CT scans, which are likely less-severe patients, led to a higher estimated sensitivity and lower specificity of the decision rules.

A second reason for imputation of the outcome is that there is more information available to accurately predict the outcome. This may be the case when there are variables measured after baseline that are related to the outcome, or when the outcome is measured repeatedly. For example, we may have assessed quality of life at 3, 6, and 12 months. If quality of life score is 50% at 3 and 12 months, it is likely 50% at 6 months as well. We note that repeated assessments of outcomes can also be assessed with mixed-effects models or state transition models.[35]

Patients with imputed outcomes can be excluded from the final analysis, while they were included in the imputation process. This has been labeled multiple-impute-and-delete (MID).[36] The final analysis is performed only on the set of patients with observed outcomes. The advantages of MID over multiple imputation may be debated.[2,37,38]

In conclusion, it is important to include the outcome as predictor in the imputation model. It is also statistically possible to impute the outcome. Whether or not to include patients with imputed outcomes in the final analysis depends on the research question: At model development, regression coefficients may generally be similar. But imputation of outcome may be beneficial for better estimation and assessment of average risk, when the outcome is non-MCAR.

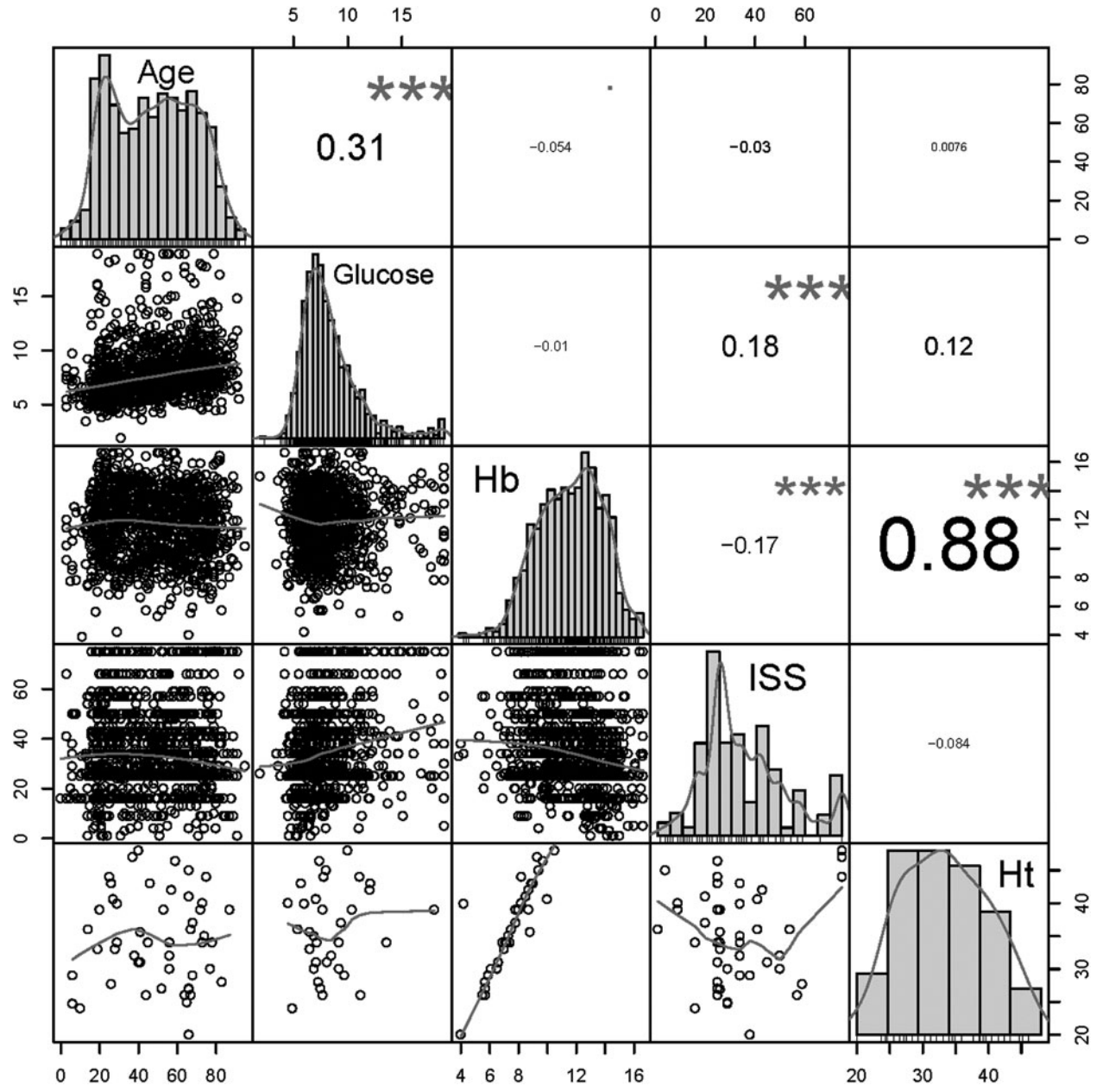

FIG. 2. Correlation plot between continuous variables, with Spearman's rank-correlation coefficients (numbers with p values: *<0.05, ** <0.01, *** <0.001). A strong correlation is observed between Hb and Ht ($r = 0.88$), whereas glucose increases somewhat with age ($r = 0.31$). Hb, hemoglobin; Ht, hematocrit; ISS, Injury Severity Score.

TABLE 4. ATTRACTIVENESS OF IMPUTATION METHOD IN PREDICTION RESEARCH FOR SCENARIOS WITH SMALL OR LARGE DATA SETS WITH SMALL OR LARGE PROPORTIONS OF MISSING DATA

Method	*Small data set, large proportion of missingness*	*Small data set, small proportion of missingness*	*Large data set, large proportion of missingness*	*Large data set, small proportion of missingness*
Complete case	–	–	–	+/–
Average imputation	–	+/–	–	+/–
Single imputation	–	+/–	+/–	+
Multiple imputation	+	+	+	+/-

Step 4: Diagnostics

After imputation, the adequacy of the imputation procedure needs to be checked. For any imputation method, it is advised to compare the distribution of imputations with the distribution of the observed values data set (Fig. 3). If distributions do not match, the question is what causes this difference. The distributions of average imputed values are different when compared to the original distribution in our case study. Distribution of multiple imputed data corresponds well with distribution of the observed variables. Moreover, convergence of the algorithm may be checked (Box 7).

Box 7: Convergence of Imputation with Chained Equations

Single and multiple imputation using the *mice* algorithm is an iterative process where missing values in multiple predictors are imputed sequentially and multiple times. This iterative process starts with a ''best guess'' of a missing value and refines this in subsequent iterations. This iterative process is finished when the algorithm that imputes missing values is converged. Whether or not convergence is reached can be studied using diagnostic plots (Supplementary Appendix SB). If these diagnostic plots show non-convergence of the algorithm, the number of iterations can be increased or the imputation model can be refined. Convergence problems can occur especially when imputing missing values for variables that are functionally dependent on each other; for instance, when imputing missing values for weight using the body mass index and, subsequently, imputing missing values for body mass index using weight.[1]

Step 5: Analysis

The final step is to use the (imputed) data set for the analysis. As stated before, we consider model development and model validation.

At model development, we fit the model of interest (Box 8) for complete case analysis, average imputation, and single imputation by conditional estimates. For multiple imputed data sets, the analyses are performed on each of the separate imputed data sets, and the results are pooled using Rubin's rules[22] (see Supplementary Appendix SD).

Box 8: Model Selection and Multiple Imputation

Model selection often precedes model estimation. Two strategies are common. First, step-wise selection may be performed with a forward selection strategy. The computer selects predictors based on a significant contribution over a smaller model. If step-wise selection is used, backward elimination is preferable. Predictors are dropped from a full model if this does not result in a significantly worse model fit. The second strategy bases selection on subject knowledge, from earlier studies and or from medical experts. This may be the most advantageous in terms of external validity[39,40]: By selecting a model with well-known predictors, the predictions become is less dependent on the specifics of the current data set at hand.

For complete case analysis and single imputation, these selection strategies can be implemented directly. For multiple imputed data sets, the last strategy (fitting a model based on earlier knowledge) can be directly implemented.

For model selection using multiple imputed data, there are several potential approaches[41]:

1. Majority: Select variables that are included in the majority of the methods.
2. Stack: Stack the imputed data sets into a single data set and use weighting to adjust for multiple occurrences of the same patient and apply the usual variable selection methods.
3. Pool and test: Perform step-wise selection based on the pooled regression coefficients and associated standard errors using the Wald test (''D1'' function in mice) or a likelihood ratio test (''D3'' function in mice).

When we re-estimated the IMPACT model, we found that most imputation methods led to similar estimates of the regression coefficients (Fig. 4). The only imputation method with markedly different estimates was the complete case analysis. Indeed, relatively few patients were in the complete case data set ($n = 589$) compared to the total cohort ($n = 1375$). Estimation of predictor effects suffers from small numbers in the complete case analysis. Interestingly, the refitted coefficients were generally more extreme (further away from 0) compared to the estimates from the IMPACT data sets.

The estimated prediction model should be validated to ensure reliable risk prediction for new patients.[42] At the developmental stage, one should at least perform some form of internal validation. A recommended method is a bootstrap procedure (Box 9).[43]

Box 9: Bootstrap Procedure for Internal Validation

The possible bootstrap validation procedure is visually displayed in Figure 5: After sampling patients with replacement, the model is fitted on that sample (the bootstrap sample) and its performance assessed in the bootstrap sample and the original data. The difference indicates the optimism in performance. This procedure is repeated many times (e.g., 200 times), and the average is taken.[44] This procedure is less straightforward after multiple imputation, given that we have multiple data sets on which we can perform bootstrapping and different approaches to pool results. Simply taking the average of the optimism to correct the average of the apparent performance over multiple imputed data sets may be reasonable. Simulation studies show that this approach produces quite valid estimates.[45]

A more rigorous test for model performance is external validation: The model is applied to a new data set of patients not used at model development. The previously fitted model (with one set of coefficients) can be applied to predict the outcome in each imputed data set. In each data set, the performance of the model can be tested and consecutively pooled. For pooling of the calibration intercept and calibration slope, Rubin's rules (Supplementary Appendix D) can be used to estimate the mean performance and the variance. For the c-statistic, the approach described above (pooling the bootstrap replicate results) can be used. These approaches were performed for validation of the IMPACT model in CENTER-TBI: It was confirmed that the complete case strategy results in different estimates for discrimination and calibration (Fig. 6). The model calibration was worse in this small subset. In contrast, discrimination

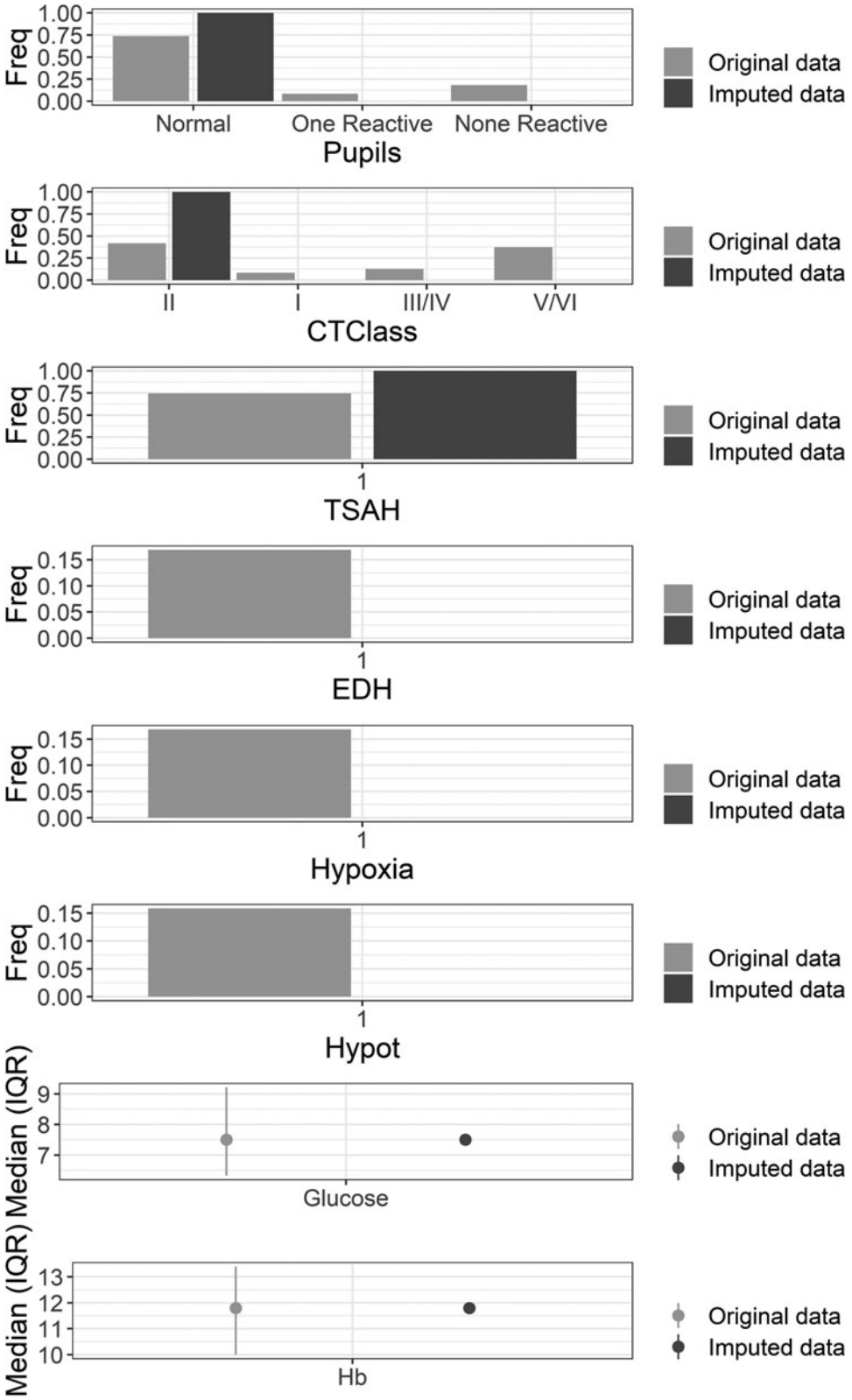

FIG. 3. Distribution of variables before and after imputation (panel **A:** average imputation; panel **B:** multiple imputation). For categorical variables, the proportion of the total number of observed values is shown, and for continuous variables, the median and interquartile range are shown. Pupils indicates pupillary reactivity. CTClass, Marshall CT class; EDH, epidural hematoma; Freq, frequency; Hb, hemoglobin; Hypot, hypotension; IQR, interquartile range; TSAH, traumatic subarachnoid hemorrhage.

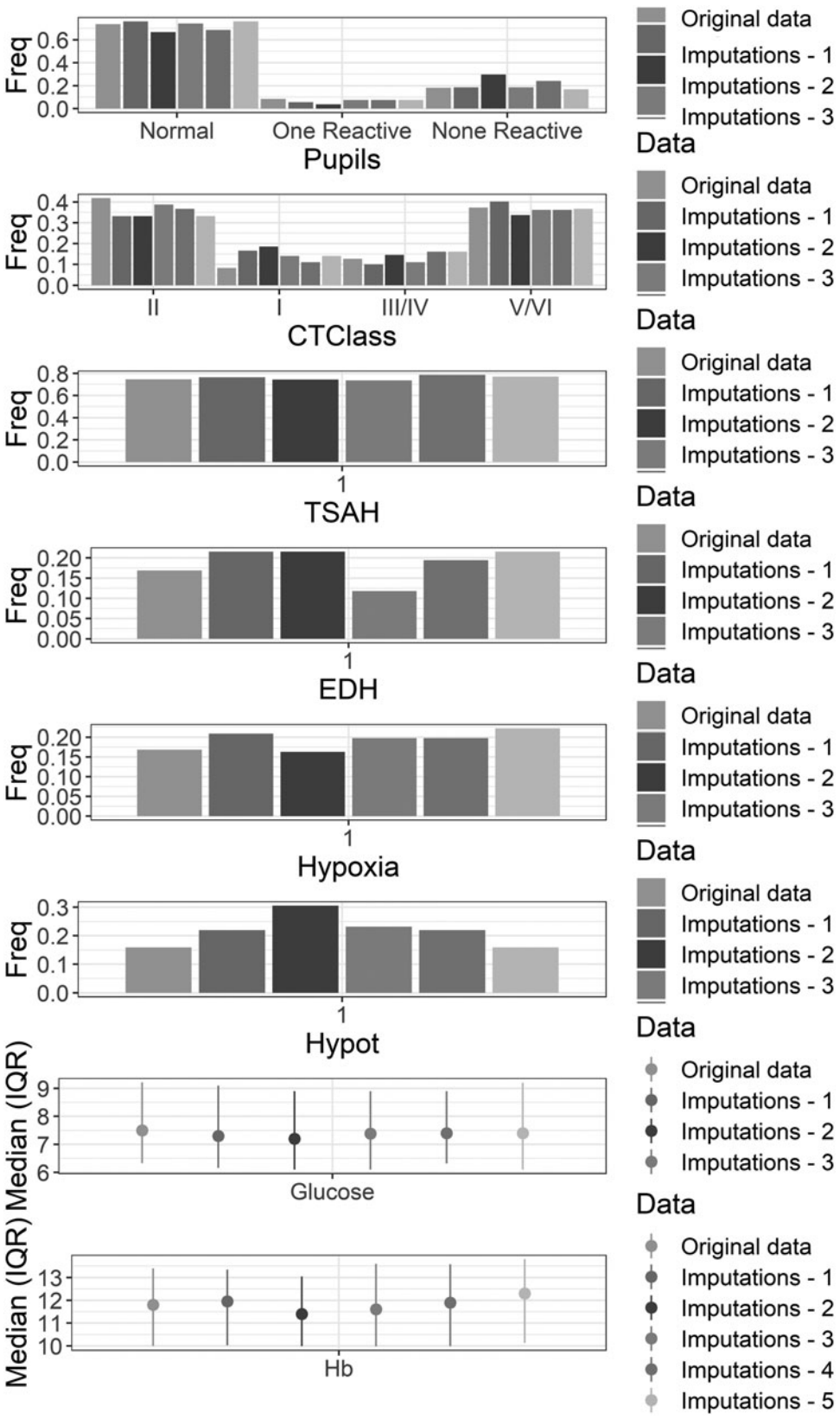

FIG. 3. (*Continued*).

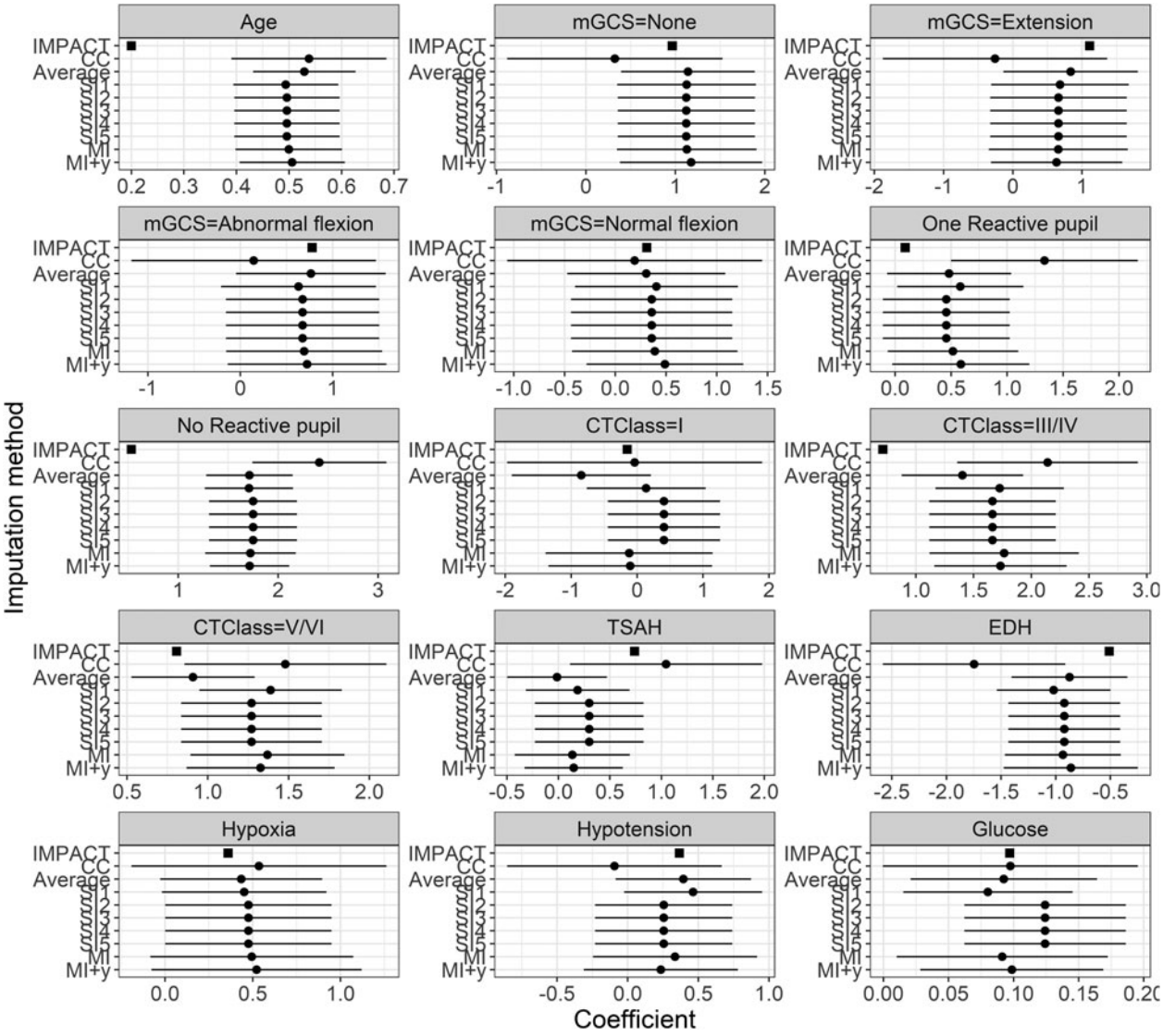

FIG. 4. Re-estimation of the IMPACT model for mortality in the completed Center-TBI data sets. The IMPACT square represents the point estimate of the IMPACT model in the original data set (18). The complete case data set includes 589 patients, the MI + y data set contains 1375 patients, whereas the other imputed data sets contain 1077 patients. CC, complete case; CTClass, Marshall CT class; EDH, epidural hematoma; IMPACT, International Mission on Prognosis and Analysis of Clinical Trials in Traumatic Brain Injury; MI, multiple imputation; MI+y, multiple imputation, also the outcome; SI, single imputation by conditional estimates; TSAH, traumatic subarachnoid hemorrhage.

was better and more uncertain. Interestingly, the multiple imputed data sets with imputed outcome (MI+y) had slightly worse calibration than the data sets where the patients with missing outcome were excluded. Interestingly, most imputations of the outcome resulted in a positive outcome (Supplementary Appendix SA, Fig. SA2).

Finally, when calibration is poor, the model can be recalibrated by re-estimating the model parameters in the new data set, and again validated. This iterative process is inherent to prediction research to ensure applicability and reliability of a prediction model.[3]

Discussion

This tutorial considered the role of imputation approaches to alleviate the problem of missing data in prediction research in acute medicine. There has been much debate about the reliability and applicability of prediction model studies, even when published in high-impact journals, because of their non-adherence to methodological principles.[46] In this context, we recommend multiple imputation, because it is a readily implementable procedure and generally superior to simple approaches, such as a complete case analysis. However, simpler imputation methods (single imputation by conditional estimates, average imputation) can be considered when the proportion of missing data is small and the database large (Table 4). Technically, it is important to include the outcome in the imputation model. For prediction of absolute risk, it may even be considered to also impute missing outcomes. Although this approach is counterintuitive, it can increase statistical efficiency and avoid biased average risk estimates. It should be noted that the guidance provided in our tutorial needs further underpinning. Specifically, more evidence from simulation studies is needed.

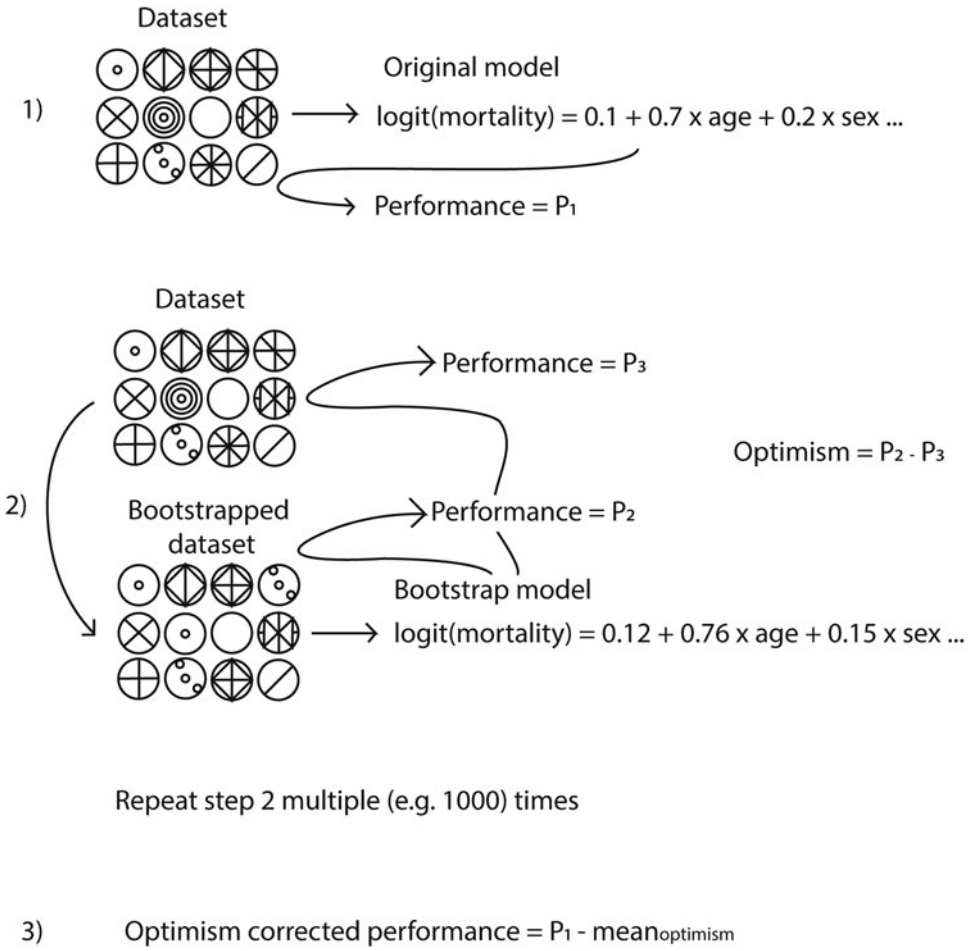

FIG. 5. Diagram showing the recommended technique for internal validation, bootstrapping. First, the model is fitted on the original data and the performance is determined in this data set. Step 2 is the bootstrap procedure: patients are drawn with replacement from the original data set to arrive at the bootstrapped data set. The model is again fitted on the bootstrapped data set. Finally, the performance in the bootstrapped data set and the original data set is obtained. The difference between performance is the optimism. Step 2 is repeated a number of times (e.g., 1000) to obtain multiple estimates of the optimism. Step 3 is to correct the originally obtained performance by the mean of the optimism obtained in step 2. IMPACT, International Mission on Prognosis and Analysis of Clinical Trials in Traumatic Brain Injury.

Other approaches

Next to imputation, there are other relevant approaches to deal with missing data. We will discuss two of these approaches.

A promising strategy to deal with missing data at external validation relies on pattern submodels.[47] Models are fitted in a development setting for each pattern of missingness. For example, one submodel might fit all predictors except for age in patients with non-observed age values, one submodel might fit all predictors except for sex and age in patients with non-observed age and sex values, and so forth. These models might be more robust, given that in contrast to imputation, they do not assume MAR.[47] Further, these models are practical in the real-world scenario: When a prediction needs to be made for a new patient, but not all predictors are measured, these submodels can still be applied. When imputation is used to develop a model, as we have advocated, either all predictors need to be collected or imputed again at validation. The sample sizes needed for validation with pattern submodels are larger than with imputation, given that a reasonable number of patients are needed for each missing data pattern.

The second other strategy to deal with missing data is inverse probability weighting.[48] In this approach, only complete cases are used, but weighted with the inverse of the probability of being a complete case. The more likely all variables are observed in a patient, the less they contribute: They are already represented well in the data set. This approach aims to mitigate selection bias associated with complete case analysis. However, because it uses only a subset of all patients, it may be less efficient than multiple imputation, which exploits the availability of information on non-missing predictors for partly complete patients.[49]

When imputation is the strategy of choice, the suggested five-step approach emphasizes key considerations and pitfalls which might be encountered. Additionally, the online vignette provides directly implementable code in the freely available R software (Supplementary Appendix C), such that the discussed approaches can easily be applied for prediction modeling in acute medicine.

The CENTER-TBI Collaborators

Cecilia Åkerlund,[1] Krisztina Amrein,[2] Nada Andelic,[3] Lasse Andreassen,[4] Audny Anke,[5] Anna Antoni,[6] Gérard Audibert,[7]

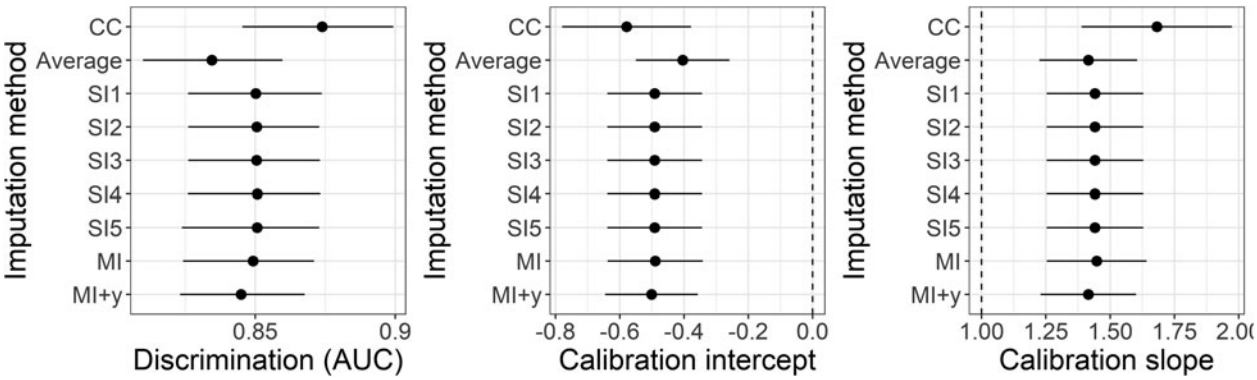

FIG. 6. External validation of the IMPACT model in variants of the CENTER TBI data set. Estimates and 95% confidence intervals are shown. AUC, area under the receiver operating characteristic curve; CC, complete case; IMPACT, International Mission on Prognosis and Analysis of Clinical Trials in Traumatic Brain Injury; MI, multiple imputation; MI+y, multiple imputation, also the outcome; SI, single imputation.

Philippe Azouvi,[8] Maria Luisa Azzolini,[9] Ronald Bartels,[10] Pál Barzó,[11] Romuald Beauvais,[12] Ronny Beer,[13] Bo-Michael Bellander,[14] Antonio Belli,[15] Habib Benali,[16] Maurizio Berardino,[17] Luigi Beretta,[9] Morten Blaabjerg,[18] Peter Bragge,[19] Alexandra Brazinova,[20] Vibeke Brinck,[21] Joanne Brooker,[22] Camilla Brorsson,[23] Andras Buki,[24] Monika Bullinger,[25] Manuel Cabeleira,[26] Alessio Caccioppola,[27] Emiliana Calappi,[27] Maria Rosa Calvi,[9] Peter Cameron,[28] Guillermo Carbayo Lozano,[29] Marco Carbonara,[27] Giorgio Chevallard,[30] Arturo Chieregato,[30] Giuseppe Citerio,[31,32] Maryse Cnossen,[33] Mark Coburn,[34] Jonathan Coles,[35] D. Jamie Cooper,[36] Marta Correia,[37] Amra Čović,[38] Nicola Curry,[39] Endre Czeiter,[24] Marek Czosnyka,[26] Claire Dahyot-Fizelier,[40] Helen Dawes,[41] Véronique De Keyser,[42] Vincent Degos,[16] Francesco Della Corte,[43] Hugo den Boogert,[10] Bart Depreitere,[44] Đula Đilvesi,[45] Abhishek Dixit,[46] Emma Donoghue,[22] Jens Dreier,[47] Guy-Loup Dulière,[48] Ari Ercole,[46] Patrick Esser,[41] Erzsébet Ezer,[49] Martin Fabricius,[50] Valery L. Feigin,[51] Kelly Foks,[52] Shirin Frisvold,[53] Alex Furmanov,[54] Pablo Gagliardo,[55] Damien Galanaud,[16] Dashiell Gantner,[28] Guoyi Gao,[56] Pradeep George,[57] Alexandre Ghuysen,[58] Lelde Giga,[59] Ben Glocker,[60] Jagoš Golubovic,[45] Pedro A. Gomez,[61] Johannes Gratz,[62] Benjamin Gravesteijn,[33] Francesca Grossi,[43] Russell L. Gruen,[63] Deepak Gupta,[64] Juanita A. Haagsma,[33] Iain Haitsma,[65] Raimund Helbok,[13] Eirik Helseth,[66] Lindsay Horton,[67] Jilske Huijben,[33] Peter J. Hutchinson,[68] Bram Jacobs,[69] Stefan Jankowski,[70] Mike Jarrett,[21] Ji-yao Jiang,[56] Kelly Jones,[51] Mladen Karan,[47] Angelos G. Kolias,[68] Erwin Kompanje,[71] Daniel Kondziella,[50] Evgenios Koraropoulos,[46] Lars-Owe Koskinen,[72] Noémi Kovács,[73] Alfonso Lagares,[61] Linda Lanyon,[57] Steven Laureys,[74] Fiona Lecky,[75] Rolf Lefering,[76] Valerie Legrand,[77] Aurelie Lejeune,[78] Leon Levi,[79] Roger Lightfoot,[80] Hester Lingsma,[33] Andrew I.R. Maas,[42] Ana M. Castaño-León,[61] Marc Maegele,[81] Marek Majdan,[20] Alex Manara,[82] Geoffrey Manley,[83] Costanza Martino,[84] Hugues Maréchal,[48] Julia Mattern,[85] Catherine McMahon,[86] Béla Melegh,[87] David Menon,[46] Tomas Menovsky,[42] Davide Mulazzi,[27] Visakh Muraleedharan,[57] Lynnette Murray,[28] Nandesh Nair,[42] Ancuta Negru,[88] David Nelson,[1] Virginia Newcombe,[46] Daan Nieboer,[33] Quentin Noirhomme,[74] József Nyirádi,[2] Otesile Olubukola,[75] Matej Oresic,[89] Fabrizio Ortolano,[27] Aarno Palotie,[90–92] Paul M. Parizel,[93] Jean-François Payen,[94] Natascha Perera,[12] Vincent Perlbarg,[16] Paolo Persona,[95] Wilco Peul,[96] Anna Piippo-Karjalainen,[97] Matti Pirinen,[90] Horia Ples,[88] Suzanne Polinder,[33] Inigo Pomposo,[29] Jussi P. Posti,[98] Louis Puybasset,[99] Andreea Radoi,[100] Arminas Ragauskas,[101] Rahul Raj,[97] Malinka Rambadagalla,[102] Ruben Real,[38] Jonathan Rhodes,[103] Sylvia Richardson,[104] Sophie Richter,[46] Samuli Ripatti,[90] Saulius Rocka,[101] Cecilie Roe,[105] Olav Roise,[106,140] Jonathan Rosand,[107] Jeffrey V. Rosenfeld,[108] Christina Rosenlund,[109] Guy Rosenthal,[54] Rolf Rossaint,[34] Sandra Rossi,[95] Daniel Rueckert,[60] Martin Rusnák,[110] Juan Sahuquillo,[100] Oliver Sakowitz,[85,111] Renan Sanchez-Porras,[111] Janos Sandor,[112] Nadine Schäfer,[76] Silke Schmidt,[113] Herbert Schoechl,[114] Guus Schoonman,[115] Rico Frederik Schou,[116] Elisabeth Schwendenwein,[6] Charlie Sewalt,[33] Toril Skandsen,[117,118] Peter Smielewski,[26] Abayomi Sorinola,[119] Emmanuel Stamatakis,[46] Simon Stanworth,[39] Ana Kowark,[34] Robert Stevens,[120] William Stewart,[121] Ewout W. Steyerberg,[33,122] Nino Stocchetti,[123] Nina Sundström,[124] Anneliese Synnot,[22,125] Riikka Takala,[126] Viktória Tamás,[119] Tomas Tamosuitis,[127] Mark Steven Taylor,[20] Braden Te Ao,[51] Olli Tenovuo,[98] Alice Theadom,[51] Matt Thomas,[82] Dick Tibboel,[128] Marjolein Timmers,[71] Christos Tolias,[129] Tony Trapani,[28] Cristina Maria Tudora,[88] Peter Vajkoczy,[130] Shirley Vallance,[28] Egils Valeinis,[59] Zoltán Vámos,[49] Gregory Van der Steen,[42] Joukje van der Naalt,[69] Jeroen T.J.M. van Dijck,[96] Thomas A. van Essen,[96] Wim Van Hecke,[131] Caroline van Heugten,[132] Dominique Van Praag,[133] Thijs Vande Vyvere,[131] Audrey Vanhaudenhuyse,[16,74] Roel P.J. van Wijk,[97] Alessia Vargiolu,[32] Emmanuel Vega,[79] Kimberley Velt,[33] Jan Verheyden,[131] Paul M. Vespa,[134] Anne Vik,[117,135] Rimantas Vilcinis,[127] Victor Volovici,[65] Nicole von Steinbüchel,[38] Daphne Voormolen,[33] Petar Vulekovic,[45] Kevin K.W. Wang,[136] Eveline Wiegers,[33] Guy Williams,[46] Lindsay Wilson,[67] Stefan Winzeck,[46] Stefan Wolf,[137] Zhihui Yang,[136] Peter Ylén,[138] Alexander Younsi,[85] Frederik A. Zeiler,[46,139] Veronika Zelinkova,[20] Agate Ziverte,[59] and Tommaso Zoerle[27]

[1]Department of Physiology and Pharmacology, Section of Perioperative Medicine and Intensive Care, Karolinska Institutet, Stockholm, Sweden

[2]János Szentágothai Research Centre, University of Pécs, Pécs, Hungary

[3]Division of Surgery and Clinical Neuroscience, Department of Physical Medicine and Rehabilitation, Oslo University Hospital and University of Oslo, Oslo, Norway

[4]Department of Neurosurgery, University Hospital Northern Norway, Tromso, Norway

[5]Department of Physical Medicine and Rehabilitation, University Hospital Northern Norway, Tromso, Norway

[6]Trauma Surgery, Medical University Vienna, Vienna, Austria
[7]Department of Anesthesiology & Intensive Care, University Hospital Nancy, Nancy, France
[8]Raymond Poincare hospital, Assistance Publique–Hopitaux de Paris, Paris, France
[9]Department of Anesthesiology & Intensive Care, S Raffaele University Hospital, Milan, Italy
[10]Department of Neurosurgery, Radboud University Medical Center, Nijmegen, The Netherlands
[11]Department of Neurosurgery, University of Szeged, Szeged, Hungary
[12]International Projects Management, ARTTIC, Munchen, Germany
[13]Department of Neurology, Neurological Intensive Care Unit, Medical University of Innsbruck, Innsbruck, Austria
[14]Department of Neurosurgery & Anesthesia & Intensive Care Medicine, Karolinska University Hospital, Stockholm, Sweden
[15]NIHR Surgical Reconstruction and Microbiology Research Centre, Birmingham, United Kingdom
[16]Anesthesie-Réanimation, Assistance Publique–Hopitaux de Paris, Paris, France
[17]Department of Anesthesia & ICU, AOU Città della Salute e della Scienza di Torino–Orthopedic and Trauma Center, Torino, Italy
[18]Department of Neurology, Odense University Hospital, Odense, Denmark
[19]BehaviourWorks Australia, Monash Sustainability Institute, Monash University, Clayton, Victoria, Australia
[20]Department of Public Health, Faculty of Health Sciences and Social Work, Trnava University, Trnava, Slovakia
[21]Quesgen Systems Inc., Burlingame, California, USA
[22]Australian & New Zealand Intensive Care Research Centre, Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia
[23]Department of Surgery and Perioperative Science, Umeå University, Umeå, Sweden
[24]Department of Neurosurgery, Medical School, University of Pécs, Hungary and Neurotrauma Research Group, János Szentágothai Research Centre, University of Pécs, Pécs, Hungary
[25]Department of Medical Psychology, Universitätsklinikum Hamburg-Eppendorf, Hamburg, Germany
[26]Brain Physics Lab, Division of Neurosurgery, Department of Clinical Neurosciences, University of Cambridge, Addenbrooke's Hospital, Cambridge, United Kingdom
[27]Neuro ICU, Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico, Milan, Italy
[28]ANZIC Research Centre, Monash University, Department of Epidemiology and Preventive Medicine, Melbourne, Victoria, Australia
[29]Department of Neurosurgery, Hospital of Cruces, Bilbao, Spain
[30]NeuroIntensive Care, Niguarda Hospital, Milan, Italy
[31]School of Medicine and Surgery, Università Milano Bicocca, Milano, Italy
[32]NeuroIntensive Care, ASST di Monza, Monza, Italy
[33]Department of Public Health, Erasmus Medical Center-University Medical Center, Rotterdam, The Netherlands
[34]Department of Anaesthesiology, University Hospital of Aachen, Aachen, Germany
[35]Department of Anesthesia & Neurointensive Care, Cambridge University Hospital NHS Foundation Trust, Cambridge, United Kingdom
[36]School of Public Health & PM, Monash University and The Alfred Hospital, Melbourne, Victoria, Australia
[37]Radiology/MRI department, MRC Cognition and Brain Sciences Unit, Cambridge, United Kingdom
[38]Institute of Medical Psychology and Medical Sociology, Universitätsmedizin Göttingen, Göttingen, Germany
[39]Oxford University Hospitals NHS Trust, Oxford, United Kingdom
[40]Intensive Care Unit, CHU Poitiers, Potiers, France
[41]Movement Science Group, Faculty of Health and Life Sciences, Oxford Brookes University, Oxford, United Kingdom
[42]Department of Neurosurgery, Antwerp University Hospital and University of Antwerp, Edegem, Belgium
[43]Department of Anesthesia & Intensive Care, Maggiore Della Carità Hospital, Novara, Italy
[44]Department of Neurosurgery, University Hospitals Leuven, Leuven, Belgium
[45]Department of Neurosurgery, Clinical centre of Vojvodina, Faculty of Medicine, University of Novi Sad, Novi Sad, Serbia
[46]Division of Anaesthesia, University of Cambridge, Addenbrooke's Hospital, Cambridge, United Kingdom
[47]Center for Stroke Research Berlin, Charité–Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany
[48]Intensive Care Unit, CHR Citadelle, Liège, Belgium
[49]Department of Anaesthesiology and Intensive Therapy, University of Pécs, Pécs, Hungary
[50]Departments of Neurology, Clinical Neurophysiology and Neuroanesthesiology, Region Hovedstaden Rigshospitalet, Copenhagen, Denmark
[51]National Institute for Stroke and Applied Neurosciences, Faculty of Health and Environmental Studies, Auckland University of Technology, Auckland, New Zealand
[52]Department of Neurology, Erasmus MC, Rotterdam, The Netherlands
[53]Department of Anesthesiology and Intensive care, University Hospital Northern Norway, Tromso, Norway
[54]Department of Neurosurgery, Hadassah-hebrew University Medical center, Jerusalem, Israel
[55]Fundación Instituto Valenciano de Neurorrehabilitación (FIVAN), Valencia, Spain
[56]Department of Neurosurgery, Shanghai Renji hospital, Shanghai Jiaotong University/School of Medicine, Shanghai, China
[57]Karolinska Institutet, INCF International Neuroinformatics Coordinating Facility, Stockholm, Sweden
[58]Emergency Department, CHU, Liège, Belgium
[59]Neurosurgery Clinic, Pauls Stradins Clinical University Hospital, Riga, Latvia
[60]Department of Computing, Imperial College London, London, United Kingdom
[61]Department of Neurosurgery, Hospital Universitario 12 de Octubre, Madrid, Spain
[62]Department of Anesthesia, Critical Care and Pain Medicine, Medical University of Vienna, Vienna, Austria
[63]College of Health and Medicine, Australian National University, Canberra, New South Wales, Australia

[64]Department of Neurosurgery, Neurosciences Centre & JPN Apex Trauma Centre, All India Institute of Medical Sciences, New Delhi, India
[65]Department of Neurosurgery, Erasmus MC, Rotterdam, The Netherlands
[66]Department of Neurosurgery, Oslo University Hospital, Oslo, Norway
[67]Division of Psychology, University of Stirling, Stirling, United Kingdom
[68]Division of Neurosurgery, Department of Clinical Neurosciences, Addenbrooke's Hospital & University of Cambridge, Cambridge, United Kingdom
[69]Department of Neurology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
[70]Neurointensive Care, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, United Kingdom
[71]Department of Intensive Care and Department of Ethics and Philosophy of Medicine, Erasmus Medical Center, Rotterdam, The Netherlands
[72]Department of Clinical Neuroscience, Neurosurgery, Umeå University, Umeå, Sweden
[73]Hungarian Brain Research Program–Grant No. KTIA_13_NAP-A-II/8, University of Pécs, Pécs, Hungary
[74]Cyclotron Research Center, University of Liège, Liège, Belgium
[75]Emergency Medicine Research in Sheffield, Health Services Research Section, School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, United Kingdom
[76]Institute of Research in Operative Medicine (IFOM), Witten/Herdecke University, Cologne, Germany
[77]VP Global Project Management CNS, ICON, Paris, France
[78]Department of Anesthesiology-Intensive Care, Lille University Hospital, Lille, France
[79]Department of Neurosurgery, Rambam Medical Center, Haifa, Israel
[80]Department of Anesthesiology & Intensive Care, University Hospitals Southhampton NHS Trust, Southhampton, United Kingdom
[81]Cologne-Merheim Medical Center (CMMC), Department of Traumatology, Orthopedic Surgery and Sportmedicine, Witten/Herdecke University, Cologne, Germany
[82]Intensive Care Unit, Southmead Hospital, Bristol, Bristol, United Kingdom
[83]Department of Neurological Surgery, University of California, San Francisco, California, USA
[84]Department of Anesthesia & Intensive Care, M. Bufalini Hospital, Cesena, Italy
[85]Department of Neurosurgery, University Hospital Heidelberg, Heidelberg, Germany
[86]Department of Neurosurgery, The Walton centre NHS Foundation Trust, Liverpool, United Kingdom
[87]Department of Medical Genetics, University of Pécs, Pécs, Hungary
[88]Department of Neurosurgery, Emergency County Hospital Timisoara, Timisoara, Romania
[89]School of Medical Sciences, Örebro University, Örebro, Sweden
[90]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland
[91]Analytic and Translational Genetics Unit, Department of Medicine; Psychiatric & Neurodevelopmental Genetics Unit, Department of Psychiatry; Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts, USA
[92]Program in Medical and Population Genetics; The Stanley Center for Psychiatric Research, The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
[93]Department of Radiology, Antwerp University Hospital and University of Antwerp, Edegem, Belgium
[94]Department of Anesthesiology & Intensive Care, University Hospital of Grenoble, Grenoble, France
[95]Department of Anesthesia & Intensive Care, Azienda Ospedaliera Università di Padova, Padova, Italy
[96]Department of Neurosurgery, Leiden University Medical Center, Leiden, The Netherlands and Department of Neurosurgery, Medical Center Haaglanden, The Hague, The Netherlands
[97]Department of Neurosurgery, Helsinki University Central Hospital, Helsinki, Finland
[98]Division of Clinical Neurosciences, Department of Neurosurgery and Turku Brain Injury Centre, Turku University Hospital and University of Turku, Turku, Finland
[99]Department of Anesthesiology and Critical Care, Pitié-Salpêtrière Teaching Hospital, Assistance Publique, Hôpitaux de Paris and University Pierre et Marie Curie, Paris, France
[100]Neurotraumatology and Neurosurgery Research Unit (UNINN), Vall d'Hebron Research Institute, Barcelona, Spain
[101]Department of Neurosurgery, Kaunas University of technology and Vilnius University, Vilnius, Lithuania
[102]Department of Neurosurgery, Rezekne Hospital, Rezekne, Latvia
[103]Department of Anaesthesia, Critical Care & Pain Medicine NHS Lothian & University of Edinburg, Edinburgh, United Kingdom
[104]Director, MRC Biostatistics Unit, Cambridge Institute of Public Health, Cambridge, United Kingdom
[105]Department of Physical Medicine and Rehabilitation, Oslo University Hospital/University of Oslo, Oslo, Norway
[106]Division of Orthopedics, Oslo University Hospital, Oslo, Norway
[107]Broad Institute, Cambridge, Massachusetts, USA; Harvard Medical School, Boston, Massachusetts, USA; and Massachusetts General Hospital, Boston, Massachusetts, USA
[108]National Trauma Research Institute, The Alfred Hospital, Monash University, Melbourne, Victoria, Australia
[109]Department of Neurosurgery, Odense University Hospital, Odense, Denmark
[110]International Neurotrauma Research Organisation, Vienna, Austria
[111]Klinik für Neurochirurgie, Klinikum Ludwigsburg, Ludwigsburg, Germany
[112]Division of Biostatistics and Epidemiology, Department of Preventive Medicine, University of Debrecen, Debrecen, Hungary
[113]Department Health and Prevention, University Greifswald, Greifswald, Germany
[114]Department of Anaesthesiology and Intensive Care, AUVA Trauma Hospital, Salzburg, Austria
[115]Department of Neurology, Elisabeth-TweeSteden Ziekenhuis, Tilburg, The Netherlands
[116]Department of Neuroanesthesia and Neurointensive Care, Odense University Hospital, Odense, Denmark
[117]Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, NTNU, Trondheim, Norway
[118]Department of Physical Medicine and Rehabilitation, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway
[119]Department of Neurosurgery, University of Pécs, Pécs, Hungary

[120]Division of Neuroscience Critical Care, John Hopkins University School of Medicine, Baltimore, Maryland, USA
[121]Department of Neuropathology, Queen Elizabeth University Hospital and University of Glasgow, Glasgow, United Kingdom
[122]Department of Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands
[123]Department of Pathophysiology and Transplantation, Milan University, and Neuroscience ICU, Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico, Milano, Italy
[124]Department of Radiation Sciences, Biomedical Engineering, Umeå University, Umeå, Sweden
[125]Cochrane Consumers and Communication Review Group, Centre for Health Communication and Participation, School of Psychology and Public Health, La Trobe University, Melbourne, Victoria, Australia
[126]Perioperative Services, Intensive Care Medicine and Pain Management, Turku University Hospital and University of Turku, Turku, Finland
[127]Department of Neurosurgery, Kaunas University of Health Sciences, Kaunas, Lithuania
[128]Intensive Care and Department of Pediatric Surgery, Erasmus Medical Center, Sophia Children's Hospital, Rotterdam, The Netherlands
[129]Department of Neurosurgery, Kings College London, London, United Kingdom
[130]Neurologie, Neurochirurgie und Psychiatrie, Charité–Universitätsmedizin Berlin, Berlin, Germany
[131]icoMetrix NV, Leuven, Belgium
[132]Movement Science Group, Faculty of Health and Life Sciences, Oxford Brookes University, Oxford, United Kingdom
[133]Psychology Department, Antwerp University Hospital, Edegem, Belgium
[134]Director of Neurocritical Care, University of California, Los Angeles, California, USA
[135]Department of Neurosurgery, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway
[136]Department of Emergency Medicine, University of Florida, Gainesville, Florida, USA
[137]Department of Neurosurgery, Charité–Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany
[138]VTT Technical Research Centre, Tampere, Finland
[139]Section of Neurosurgery, Department of Surgery, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Manitoba, Canada
[140]Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway

Funding Information

Data used in preparation of this article were obtained in the context of CENTER-TBI, a large collaborative project with the support of the European Union 7th Framework program (EC grant 602150). Additional funding was obtained from the Hannelore Kohl Stiftung (Germany), the OneMind (USA), and the Integra LifeSciences Corporation (USA). The funder had no role in the study design, enrollment, collection of data, writing, or publication decisions.

Author Disclosure Statement

No competing financial interests exist.

Supplementary Material

Supplementary Appendix SA
Supplementary Appendix SB
Supplementary Appendix SC
Supplementary Appendix SD
Supplementary Appendix SA Table SA1
Supplementary Appendix SA Figure SA1
Supplementary Appendix SA FIgure SA1

References

1. van Buuren, S. (2018). *Flexible Imputation of Missing Data*. CRC: Boca Raton, FL.
2. Steyerberg, E.W. (2019). *Clinical Prediction Models*. Springer International: Cham, Switzerland.
3. Steyerberg, E., Moons, K.G.M., van der Windt, D., Hayden, J., Perel, P., Schroter, S., Riley, R., Hemingway, H., and Altman, R.B. (2013). Prognosis Research Strategy (PROGRESS) series 3: prognostic model research. PLoS Med. 10, e1001381.
4. Riley, R.D., Hayden, J.A., Steyerberg, E.W., Moons, K.G.M., Abrams, K., Kyzas, P.A., Malats, N., Briggs, A., Schroter, S., Altman, D.G., and Hemingway, H. (2013). Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. PLoS Med. 10, e1001380.
5. Molenberghs, G., Kenward, M., and Ebrahim, G.J. (2007). Missing data in clinical studies. J. Trop. Pediatr. 53, 294–294.
6. Mislevy, R.J. (1991). Book reviews: statistical analysis with missing data. J. Educ. Stat. 16, 150.
7. Nielson, J.L., Cooper, S.R., Seabury, S.A., Luciani, D., Fabio, A., Temkin, N.R., and Ferguson, A.R.; TRACK-TBI Investigators. (2020). Statistical guidelines for handling missing data in traumatic brain injury clinical research. J. Neurotrauma. doi: 10.1089/neu.2019.6702.
8. Leisman, D.E., Harhay, M.O., Lederer, D.J., Abramson, M., Adjei, A.A., Bakker, J., Ballas, Z.K., Barreiro, E., Bell, S.C., Bellomo, R., Bernstein, J.A., Branson, R.D., Brusasco, V., Chalmers, J.D., Chokroverty, S., Citerio, G., Collop, N.A., Cooke, C.R., Crapo, J.D., Donaldson, G., Fitzgerald, D.A., Grainger, E., Hale, L., Herth, F.J., Kochanek, P.M., Marks, G., Moorman, J.R., Ost, D.E., Schatz, M., Sheikh, A., Smyth, A.R., Stewart, I., Stewart, P.W., Swenson, E.R., Szymusiak, R., Teboul, J.-L., Vincent, J.-L., Wedzicha, J.A., and Maslove, D.M. (2020). Development and reporting of prediction models: guidance for authors from editors of respiratory, sleep, and critical care journals. Crit. Care Med. 48, 623–633.
9. Richter, S., Stevenson, S., Newman, T., Wilson, L., Menon, D., Maas, A., Nieboer, D., Lingsma, H., Steyerberg, E.W., and Newcombe, V. (2019). Handling of missing outcome data in traumatic brain injury research: a systematic review. J. Neurotrauma 36, 2743–2752.
10. Maas, A.I.R., Menon, D.K., Steyerberg, E.W., Citerio, G., Lecky, F., Manley, G.T., Hill, S., Legrand, V., and Sorgner, A.; CENTER-TBI Participants and Investigators. (2015). Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI): a prospective longitudinal observational study. Neurosurgery 76, 67–80.
11. Steyerberg, E.W., Wiegers, E., Sewalt, C., Buki, A., Citerio, G., Keyser, V. De, Ercole, A., and Kunzmann, K. (2019). Case-mix , care pathways , and outcomes in patients with traumatic brain injury in CENTER-TBI: a European prospective, multicentre, longitudinal, cohort study. Lancet Neurol. 18, 923–934.
12. Shmueli, G. (2010). To explain or to predict? Stat. Sci. 25, 289–310.
13. Steyerberg, E.W., Mushkudiani, N., Perel, P., Butcher, I., Lu, J., McHugh, G.S., Murray, G.D., Marmarou, A., Roberts, I., Habbema, J.D.F., and Maas, A.I.R. (2008). Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. PLoS Med. 5, e165; discussion, e165.
14. Perel, P., Arango, M., Clayton, T., Edwards, P., Komolafe, E., Poccock, S., Roberts, I., Shakur, H., Steyerberg, E., and Yutthakasemsunt, S. (2008). Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients. BMJ 336, 425–429.
15. Dijkland, S.A., Foks, K.A., Polinder, S., Dippel, D.W.J., Maas, A.I.R., Lingsma, H.F., and Steyerberg, E.W. (2020). Prognosis in moderate and severe traumatic brain injury: a systematic review of contemporary models and validation studies. J. Neurotrauma 37, 1–13.

16. MRC CRASH Trial Collaborators. (2008). Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients. BMJ 336, 425–429.
17. Cox, D. (1958). Two further applications of a model for binary regression. Biometrika 45, 562–565.
18. White, I.R., and Carlin, J.B. (2010). Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. Stat. Med. 29, 2920–2931.
19. Collins, G.S., Reitsma, J.B., Altman, D.G., and Moons, K.G.M. (2015). Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann. Intern. Med. 162, 55–63.
20. Hughes, R.A., Heron, J., Sterne, J.A.C., and Tilling, K. (2019). Accounting for missing data in statistical analyses: Multiple imputation is not always the answer. Int. J. Epidemiol. 48, 1294–1304.
21. Van Houwelingen, H.C., and Sauerbrei, W. (2013). Cross-validation, shrinkage and variable selection in linear regression revisited. Open J. Stat. 3, 79–102.
22. Rubin, D.B. (2004). *Multiple Imputation for Nonresponse in Surveys*. Wiley-Interscience: Hoboken, NJ.
23. Lingsma, H.F., Roozenbeek, B., Steyerberg, E.W., Murray, G.D., and Maas, A.I. (2010). Early prognosis in traumatic brain injury: from prophecies to predictions. Lancet Neurol. 9, 543–554.
24. Roozenbeek, B., Lingsma, H.F., Lecky, F.E., Lu, J., Weir, J., Butcher, I., McHugh, G.S., Murray, G.D., Perel, P., Maas, A.I., and Steyerberg, E.W.; International Mission on Prognosis Analysis of Clinical Trials in Traumatic Brain Injury (IMPACT) Study Group, Corticosteroid Randomisation After Significant Head Injury (CRASH) Trial Collaborators, and Trauma Audit and Research Network (TARN). (2012). Prediction of outcome after moderate and severe traumatic brain injury. Crit. Care Med. 40, 1609–1617.
25. Majdan, M., Lingsma, H.F., Nieboer, D., Mauritz, W., Rusnak, M., and Steyerberg, E.W. (2014). Performance of IMPACT, CRASH and Nijmegen models in predicting six month outcome of patients with severe or moderate TBI: an external validation study. Scand. J. Trauma. Resusc. Emerg. Med. 22, 68.
26. Lingsma, H., Andriessen, T.M.J.C., Haitsema, I., Horn, J., van der Naalt, J., Franschman, G., Maas, A.I.R., Vos, P.E., and Steyerberg, E.W. (2013). Prognosis in moderate and severe traumatic brain injury. J. Trauma Acute Care Surg. 74, 639–646.
27. Teasdale, G., and Jennett, B. (1974). Assessment of coma and impaired consciousness. A practical scale. Lancet 2, 81–84.
28. Steyerberg, E.W., Nieboer, D., Debray, T.P.A., and Houwelingen, H.C. (2019). Assessment of heterogeneity in an individual participant data meta-analysis of prediction models: an overview and illustration. Stat. Med. 38, 4290–4309.
29. Lee, K.J., and Carlin, J.B. (2012). Recovery of information from multiple imputation: a simulation study. Emerg. Themes Epidemiol. 9, 3.
30. Madley-Dowd, P., Hughes, R., Tilling, K., and Heron, J. (2019). The proportion of missing data should not be used to guide decisions on multiple imputation. J. Clin. Epidemiol. 110, 63–73.
31. Von Hippel, P., and Lynch, J. (2013). Efficiency gains from using auxiliary variables in imputation. arxiv.org/ftp/arxiv/papers/1311/1311.5249.pdf (Last accessed February 3, 2021).
32. Mustillo, S. (2012). The effects of auxiliary variables on coefficient bias and efficiency in multiple imputation. Sociol. Methods Res. 41, 335–361.
33. Moons, K.G.M., Donders, R.A.R.T., Stijnen, T., and Harrell, F.E. (2006). Using the outcome for imputation of missing predictor values was preferred. J. Clin. Epidemiol. 59, 1092–1101.
34. Foks, K.A., Van Den Brand, C.L., Lingsma, H.F., Van Der Naalt, J., Jacobs, B., De Jong, E., Den Boogert, H.F., Sir, Ö., Patka, P., Polinder, S., Gaakeer, M.I., Schutte, C.E., Jie, K.E., Visee, H.F., Hunink, M.G.M., Reijners, E., Braaksma, M., Schoonman, G.G., Steyerberg, E.W., Jellema, K., and Dippel, D.W.J. (2018). External validation of computed tomography decision rules for minor head injury: prospective, multicentre cohort study in the Netherlands. BMJ 362, k3527.
35. Kunzmann, K., Wernisch, L., Richardson, S., Steyerberg, E., Lingsma, H., Ercole, A., Maas, A., Menon, D., and Wilson, L. (2021). Imputation of ordinal outcomes: a comparison of approaches in traumatic brain injury. J. Neurotrauma. 38, 455–463.
36. von Hippel, P.T. (2007). Regression with missing Ys: an improved strategy for analyzing multiply imputed data. Sociol. Methodol. 37, 83–117.
37. Kontopantelis, E., White, I.R., Sperrin, M., and Buchan, I. (2017). Outcome-sensitive multiple imputation: a simulation study. BMC Med. Res. Methodol. 17, 2.
38. Sullivan, T.R., Salter, A.B., Ryan, P., and Lee, K.J. (2015). Bias and precision of the ''multiple imputation, then deletion'' method for dealing with missing outcome data. Am. J. Epidemiol. 182, 528–534.
39. Wahlby, U., Jonsson, E.N., and Karlsson, M.O. (2002). Comparison of stepwise covariate model building strategies in population pharmacokinetic-pharmacodynamic analysis. AAPS PharmSci. 4, E27.
40. Derksen, S., and Keselman, H.J. (1992). Backward, forward and stepwise automated subset selection algorithms: frequency of obtaining authentic and noise variables. Br. J. Math. Stat. Psychol. 45, 265–282.
41. Wood, A.M., White, I.R., and Royston, P. (2008). How should variable selection be performed with multiply imputed data? Stat. Med. 27, 3227–3246.
42. Steyerberg, E.W., and Harrell, F.E. (2016). Prediction models need appropriate internal, internal-external, and external validation HHS Public Access. J. Clin. Epidemiol. 69, 245–247.
43. Steyerberg, E.W., Harrell, F.E., Borsboom, G.J.J.M., Eijkemans, M.J.C., Vergouwe, Y., and Habbema, J.D.F. (2001). Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J. Clin. Epidemiol. 54, 774–781.
44. Efron, B., and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. CRC: Boca Raton, FL.
45. Schomaker, M., and Heumann, C. (2018). Bootstrap inference when using multiple imputation. Stat. Med. 37, 2252–2266.
46. Bouwmeester, W., Zuithoff, N.P.A., Mallett, S., Geerlings, M.I., Vergouwe, Y., Steyerberg, E.W., Altman, D.G., and Moons, K.G.M. (2012). Reporting and methods in clinical prediction research: a systematic review. PLoS Med. 9, e1001221.
47. Mercaldo, S.F., and Blume, J.D. (2020). Missing data and prediction: the pattern submodel. Biostatistics 21, 236–252.
48. Seaman, S.R., and White, I.R. (2013). Review of inverse probability weighting for dealing with missing data. Stat. Methods Med. Res. 22, 278–295.
49. Perkins, N.J., Cole, S.R., Harel, O., Tchetgen Tchetgen, E.J., Sun, B., Mitchell, E.M., and Schisterman, E.F. (2018). Principled approaches to missing data in epidemiologic studies. Am. J. Epidemiol. 187, 568–575.

Address correspondence to:
Benjamin Yaël Gravesteijn, MD
Department of Public Health
Erasmus Medical Center
Postbus 2040
3000 CA
Rotterdam
The Netherlands

E-mail: b.gravesteijn@erasmusmc.nl