



Urko Aguirre-Larracoechea ^{1,2,3,*} and Cruz E. Borges ⁴

- ¹ Research Unit, Osakidetza Basque Health Service, Barrualde-Galdakao Integrated Health Organisation, Galdakao-Usansolo Hospital, 48960 Galdakao, Spain
- ² Kronikgune Institute for Health Services Research, 48902 Barakaldo, Spain
- ³ Research Network in Health Services in Chronic Diseases (Red de Investigación en Servicios de Salud en Enfermedades Crónicas, REDISSEC), 48960 Galdakao, Spain
- ⁴ Deusto Institute of Technology, Faculty of Engineering, University of Deusto, 48007 Bilbao, Spain; cruz.borges@deusto.es
- * Correspondence: urko.aguirrelarracoechea@osakidetza.eus

Abstract: Real-life data are bounded and heavy-tailed variables. Zero-one-inflated beta (ZOIB) regression is used for modelling them. There are no appropriate methods to address the problem of missing data in repeated bounded outcomes. We developed an imputation method using ZOIB (i-ZOIB) and compared its performance with those of the naïve and machine-learning methods, using different distribution shapes and settings designed in the simulation study. The performance was measured employing the absolute error (MAE), root-mean-square-error (RMSE) and the unscaled mean bounded relative absolute error (UMBRAE) methods. The results varied depending on the missingness rate and mechanism. The i-ZOIB and the machine-learning ANN, SVR and RF methods showed the best performance.

Keywords: imputation; bounded outcomes; repeated measures; zero-one-inflated beta distribution; machine learning

1. Introduction

In any research study, one of the most important tasks is data analysis. The results of such analysis can support or refute the hypotheses proposed by the researchers. It is, therefore, important to have high-quality data to draw and extrapolate the conclusions. Several different aspects can be considered in the quality assessment of the data, such as time necessary to ensure that the value is new, consistency (the representation of the data should be invariant in all cases), completeness (the data should be complete, without missing values), and accuracy (to ensure that the recorded value is identical to the true value) [1].

Missing data are among the most frequent and often-evaluated problems in all types of surveys, especially in repeated or longitudinal studies. In the latter type of studies, the missingness or dropout rates can be affected by many known and unrelated factors such as refusal to participate, death of the subject, etc. Little and Rubin [2] have classified these factors into three groups: (a) missing completely at random (MCAR): the missing information is due to chance; (b) missing at random (MAR): the lack of information is conditioned solely by the observed values; and finally, (c) missing not at random (MNAR): the missing information depends on both missing and non-missing information.

Numerous studies conclude that imputing unobserved data is better than ignoring them [3–5]. The most used classical imputation methods replace the unobserved value by the mean or median, by the last observed value, or use multiple imputation or regression [6–9]. However, in most cases, the response variable is assumed to be an unbounded real value that follows a normal distribution. In real life, and increasingly



Citation: Aguirre-Larracoechea, U.; Borges, C.E. Imputation for Repeated Bounded Outcome Data: Statistical and Machine-Learning Approaches. *Mathematics* **2021**, *9*, 2081. https:// doi.org/10.3390/math9172081

Academic Editors: María Del Mar Rueda, Andrés Cabrera-León and Alfonso Mateos Caballero

Received: 7 July 2021 Accepted: 25 August 2021 Published: 28 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). in many disciplines such as medicine, biology, psychology, or engineering, we are faced with situations in which the data used to make predictions are variables that are either strictly positive or bounded in an interval (a,b). The most relevant examples are the variables holding length or volume (strictly positive) or obtained by grading questionnaires or clinical scores. Moreover, by mathematical transformation, any value in the interval (a,b) can be univocally identified with another value in the interval (0,1) [10,11]. One of the recommended statistical techniques addressing this problem is beta regression [10,12]. Nevertheless, the values of response variables might also be exactly 0 or 1; the bounding interval is closed or partially closed (namely, it contains one or both extreme values). This leads to a combination of distributions such as zero-inflated, one-inflated, or zero-one-inflated. Though the beta distribution covers various distribution shapes, it does not accommodate extreme values at 0 and 1. There are very few studies using the zero-one-inflated beta regression for the imputation of missing data.

In the last decade, the use of artificial intelligence (AI) and machine-learning (ML) techniques has increased substantially in a diverse range of disciplines, such as climatology, industry, and biomedicine [13–20]. There is a vast and diverse number of ML algorithms which may be mostly categorized based on the approach taken toward the data set as well as the type of data processed. For instance, the well-known K-Nearest Neighbours (K-NN) and support vector machines (SVM) models are considered as classical ML techniques. Furthermore, newer generation AI and ML techniques that provide higher performance than the classical tools have been developed: the artificial neural networks (ANN), random forest (RF), extreme Gradient Boosting (XGBOOST) and light gradient boosting models (LightGBoost).

These abovementioned techniques (K-NN, SVM, ANN, Random Forest, and Gradient Boost Machine based models) were also developed and used to investigate their performance when handling missing data. Aguilera et al. [21] used the RF models, the predictive mean matching and the spatio-temporal krigging as powerful tools for the missing data imputation within the hydrological spatio-temporal setting. They concluded that the latter models were the most robust method. To establish another approach for handling missing data, Asraf H et al. [22] deployed several ML and AI strategies to impute missing data within the industrial environment. They concluded that support vector regression (SVR) models outperformed ANN and K-NN methods. In the biomedicine field, Zhang et al. [23] used the the XGBoost model to address the missingness issue. Thus, the review of the published works shows that the missing data imputation can be also achieved by using AI-ML-driven techniques.

To date, several papers tackling the missingness problem in the response variable or the covariates (in longitudinal or other settings) have employed statistical or artificial intelligence methods [19,24,25]. Nevertheless, these studies have assumed that the distribution of the response variable is normal. Moreover, few papers address the missingness issue for non-Gaussian response variables and even fewer for bounded dependent variables. The existing studies mostly focus on providing or comparing methodologies based on multiple imputation [26,27]. However, our interest lies in obtaining unique rather than multiple imputed distributions in repeated measures setting.

Thus, motivated from the relevance of the nature of the outcome variable [11,28], this study focuses on evaluating the performance of statistical and machine-learning methods in the imputation of response variables bounded in the closed interval [0,1] under a repeated-measures scenario. From wide-ranging review of previous works, seven different statistical-AI methods were selected: zero-one inflated beta regression (statistical approach), support vector regression (SVR), 1-Nearest Neighbour, ANN (AI nonlinear algorithm), random forest (RF), extreme gradient boosting (XGBOOST) and light gradient boosting models (LightGBoost).

The paper is divided into several sections. Section 2 describes the theoretical basis of the techniques to be compared. Section 3 presents the simulation study carried out and the corresponding results. Section 4 deals with the results obtained by the application of

the methods to the CARESS-CCR study. Finally, Section 5 discusses the results and draws some general conclusions.

2. Materials and Methods

2.1. Distributions

Beta regression is an increasingly popular statistical technique used to predict outcomes with the values in the range of (0,1), such as rates and proportions, patient-reported outcomes, or economics studies [11,29]. Nevertheless, in certain situations, outcome variables also take the extreme values. In this case, zero-or-one-inflated beta (ZOIB) regression can be employed, when outcomes have values in the range [0, 1), [0, 1), or [0, 1].

Variables bounded in the interval [0,1] can take the following forms (Figure 1):





The general idea is to model the response variable (call it y) as a mixture of Bernoulli and beta distributions, from which the true 0 s and 1 s and the values between 0 and 1 are generated. The probability density function is:

$$f_{ZOIB}(y; \alpha, \gamma, \mu, \phi) = \begin{cases} \alpha(1-\gamma) & y=0\\ \alpha\gamma & y=1\\ (1-\alpha)f(y; \mu; \phi) & 0 < y < 1 \end{cases}$$
(1)

where $0 < \alpha$, γ , $\mu < 1$ and $\phi > 0$. $f(y; \mu, \phi)$ is the probability density function for the beta distribution, parameterised in terms of its mean μ and precision ϕ :

$$f_{beta}(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1} (1-y)^{(1-\mu)\phi-1}$$
(2)

 α is a mixture parameter that determines the extent to which the Bernoulli or beta component dominates the probability density function. γ determines the probability that y = 1 if it comes from the Bernoulli component. μ and ϕ are the expected value and the precision for the beta component, which is usually parameterised in terms of p and q: $\mu = p + q$ and $\phi = p + q$. [12].

It should be noted that these results can be generalised univocally to *y* variables whose values lie in the interval [a,b] by using the following mathematical transformation:

$$\mathbf{y}'' = \frac{(\mathbf{y} - \mathbf{a})}{(\mathbf{b} - \mathbf{a})} \tag{3}$$

where y'' values range within the [0, 1] interval.

2.2. Approaches for Handling Missing Data

2.2.1. Imputation via Zero-One-Inflated Regression (i-ZOIB)

In this work, we assess the performance of zero-one-inflated beta regression (ZOIB) as a missing-data imputation technique. ZOIB, based on the Bayesian regression, has been developed [11] to model-bounded outcomes whose values are within the zero-one interval. After obtaining the computed predictive values, we used the predictive mean matching methodology [7,30] to impute the missing values.

2.2.2. Artificial Intelligence (AI)—Machine-Learning (ML) Methods

K-nearest neighbour (K-NN)

The k-nearest neighbour method belongs to the machine-learning family. This approach is a donor-based method where the replaced value is either a value measured for another record, or the average of k (k > 1) measured values. In contrast to classical techniques, k-NN can be applied to mixed-continuous and categorical data. In our case, the 1-NN technique has been applied according to the literature [31].

Support vector regression

A support vector machine (SVM) has originated in the research on the theory of statistical learning; it was introduced in the 1990s by Vapnik and his collaborators [32,33]. Support vector regression (SVR) is a variant of the SVM analysis, used as a regression scheme to predict values.

The SVM represents the sample points in space, separating the classes into two regions, as wide as possible, by employing a separation hyperplane, giving rise to support vectors (formed by the data points close to the hyperplane boundary).

When a new sample is added to the SVM model, it is categorised into one of the classes, depending on the space to which it belongs. A good separation between the classes allows a correct classification. Thus, an SVM builds a hyperplane or a set of hyperplanes in the space of very high or even infinite dimensionality that can be used in classification or regression problems. To extend such a concept to non-linear decision surfaces, the kernel functions are applied to transform the original data to map into a new space. In this way, the points of the vector that are on one side of the hyperplane are labelled as belonging to one category and those on the other side to the other. The predictor variable is called an attribute, and the attributes used to define the hyperplane are called characteristics.

The input data are viewed as a p-dimensional vector (an ordered list of p numbers), as in most supervised classification methods. The SVM looks for a hyperplane that optimally separates the points of one class from another, which could have been previously projected in the space of higher dimensionality.

Support vector regression uses the same principle with some minor changes. As the output is a real number, it becomes difficult to predict the information "manually" since there are infinite possibilities. In the case of regression, a tolerance margin (epsilon) is established near the vector to minimise the error, considering that a part of that error is tolerated [34].

Random Forest (RF)

Random forests (RF) is an ensemble-learning algorithm based on decision trees and developed by Breiman [35].

Given a set of original data values $\mathbf{S} = \{(x_i, y_i)\}_{i=1}^n$ and a new independent test case C_{test} with a predictor x_{test} , the steps for developing a RF model are the next ones:

- i. The set of values S is sampled with replacement using the bootstrap technique: $B_{1, \dots, B_{L}}$
- ii. On each B_j resample (j = 1, ..., L), we grow a regression tree T_j .
- iii. Finally, the predicted value of the predictor x_{test} of the test case C_{test} by RF, is obtained as the mean average of the results given by each of the individual T_j trees. Mathematically, if y_j^* (j = 1, ..., L) is the prediction of C_{test} obtained on each of the T_j trees, then the final RF prediction y_{test} is reflected as follows:

$$y_{test} = rac{1}{L}\sum_{j=1}^{L}y_j^*$$

RF is generally used for both data classification and regression. For its performance, RF needs to fix the number of bootstrap samples (ntree), number of predictors (mtry) and the number of threes (ntree). RF hardly overfits when more trees are added, but produce a limited generalisation error [36].

Artificial Neural Network (ANN)

An ANN is an algorithm based on a nervous system analogy. The general idea consists of emulating the learning capacity of the nervous system. Figure 2 shows a graphical illustration of a basic ANN architecture. The ANN learns to identify a pattern of association between the values of a set of variables predictors $({X_i}_{i=1}^n \text{ inputs})$ and the states or values that are considered dependent on these values (*Y* output). Thus, the ANN consists of a group of units of process (nodes) that resemble neurons by being interconnected through a network of relationships $({w_i}_{i=1}^n \text{ weights})$, which is represented as $z = \sum_{i=1}^n X_i \cdot w_i$. Before being transferred this information to the next node, an activation function *f* is applied to *z*. After that, the output $\alpha = f(z)$ will be moved to the subsequent node.

In the nodes, three types of operations are performed to determine the output of the neuron that are: propagation rule, activation function and output function. There are several activation functions such as the linear, ReLU, or softplus transformations to transfer the information from the hidden-layer (internal node) to the final output.



Figure 2. The basic Artificial Neural Network architecture.

Gradient Boosting algorithms: Extreme Gradient Boosting

The XGBoost (Extreme Gradient Boosting) algorithm is a supervised learning technique [37] also based on decision trees and which is considered the state of the art in the evolution of these algorithms.

It consists of a sequential assembly of decision trees. The trees are added sequentially in order to learn from the results of the previous trees and correct the error produced by them, until the error can no longer be corrected (this is known as a "downward gradient". It works as follows:

- i. An initial tree F_0 is obtained to predict the target variable *y*, the result is associated with a residual (*y*- F_0).
- ii. A new tree h_1 is obtained that adjusts to the error of the previous step.
- iii. The results of F_0 and h_1 are combined to obtain the F_1 tree, where the root mean square error of F_1 will be less than that of F_0 :

$$F_1(x) \leftarrow F_0(x) + h_1(x)$$

iv. This process is continued iteratively until the error is minimized as much as possible in the following way:

$$F_m(x) \leftarrow F_{m-1}(x) + h_m(x)$$

The main advantages of the XGBoost algorithm are: (a) it can handle large databases with multiple variables as well as missing variables; (b) excellent execution speed. The main difference between the XGBoost and Random Forest algorithms is that in the former the user defines the extent of the trees, while in the latter the trees grow to their maximum extent. It uses parallel processing, tree pruning, handling of missing values and regularization (optimization that penalizes the complexity of the models) to avoid overfitting or bias of the model as much as possible. • Gradient Boosting algorithms: Light Gradient Boosting models

This algorithm uses a tree-based learning algorithm developed by Guolin et al. [38] The biggest difference with XGBoost is that LGBM does not grow line by line at tree-level-wise, but instead at leaf-wise level, that makes the algorithm faster.

3. Simulation Study

3.1. Simulation Design

A simulation study was conducted to assess the performance of different statistical and machine-learning methods of missing data handling.

For repeated data, let us consider 500 subjects and two measurements for each subject. The functional form of the model is as follows:

$$Y_{ij} = \beta_o + \beta_1 X_{ij} + \beta_2 T_{ij} + \beta_3 X_{ij} * T_{ij} + b_i + \epsilon_{ij}$$

$$\tag{4}$$

where i = 1, ..., 500 and j = 1, 2.

For fixed effects, we constructed a design matrix $X = [X_1 \dots X_{1000}]$, where each X_k is a vector of 2 elements. The first element was constant and equal to 1, and the other elements were sampled from a beta distribution. For random effects, we constructed a design matrix $b = [b_1 \dots b_{500}]$, where each b_k is a vector of a Chi-square distribution with 3 degrees of freedom. X is fully observed.

As noted in the introduction, we focused on the management of missing data in repeated bounded outcomes. As bounded outcomes can adopt different shapes, we considered three distinct scenarios based on the designs mentioned above, which are related to real-life situations:

Scenario 1: inverse-J-shaped distribution both at the baseline and at the follow-up

In this scenario, the subjects have low scores at the beginning and during the followup.

Scenario 2: inverse-J-shaped at the baseline and inverse U-shaped at the follow-up

Here, subjects present low values at the beginning, but the values increase slightly during the follow-up.

Scenario 3: J-shaped distribution both at the baseline and at follow-up

In this scenario, the participants show high scores at both time points.

For each of the three scenarios, we generated missing observations for the followup measurement, according to the missingness mechanisms (MCAR, MAR or MNAR) and different missingness rates (10%, 20% and 30%). Given a dataset, the missingness probability p_i on Y was defined as follows:

$$p_{i} = \begin{cases} \alpha & if missingness mechanism is MCAR\\ \alpha + \beta X & if missingness mechanism is MAR\\ \alpha + \beta X + \gamma Y & if missingness mechanism is MNAR \end{cases}$$
(5)

Finally, we run this experiment 250 times with different missing value mechanisms and rates for each assessed scenario.

3.2. Setting up Parameters for Statistical and Machine-Learning Approaches

We used the statistical and machine-learning methods under each of the three scenarios. A value of k = 1 for the K-NN imputation approach has been chose, as mentioned in the literature [31]. With regard to the rest of the ML/AI approaches (SVR, RF, ANN, XGBOOST and LightGBoost), the hyperparameter tuning was carried out to determine the best optimized model performed on each of the simulated dataset and run times.

8 of 27

3.3. Accuracy Measures

Different accuracy measures were used to evaluate the performance of statistical and machine-learning imputation methods.

Given a dataset with n observations, let Y denote the true distribution of the variable of interest and Y*, the imputed variable after using each tested method. Then, we define the error $e = (Y - Y^*)$ as the difference between the true and the imputed distributions by some benchmark method. The error in the observations in which the missing data were generated was $e_{miss} = (Y - Y^*)_{miss}$. Thus, to assess the impact of the imputation technique, we only focused on those observations that had been generated missing data and imputed afterwards.

There are many accuracy measures to assess the performance [39]; we employed the most commonly used:

(a) Mean absolute error method (MAE, defined as the arithmetic mean of absolute errors).

$$MAE = \frac{1}{n_{miss}} \sum_{i=1}^{n_{miss}} |e_{miss}|$$
(6)

(b) Mean squared error (MSE)/root mean squared error (RMSE) method. The arithmetic mean of squared errors is computed. Outliers affect the MSE since they give large weight to large errors. Moreover, as the errors are squared, the results are given on a distinct scale. Thus, the RMSE (the square root of MSE) is preferable to MSE. Nevertheless, RMSE is also sensitive to outliers.

$$RMSE = \sqrt{\frac{1}{n_{miss}} \sum_{i=1}^{n_{miss}} e_{miss}^2}$$
(7)

As already noted, the MAE, MSE, and RMSE are scale-dependent accuracy measures with various advantages and disadvantages. However, it is recommended to use a measure that satisfies several related requirements, i.e., is informative, resistant to outliers, symmetric, scale-independent and easy to understand (provides intuitive results). A relatively new accuracy metric, the unscaled mean bounded relative absolute error UMBRAE [40], has been defined to tackle some of the existing problems. We included it in our imputation performance assessment. It is computed as follows:

$$MBRAE = \frac{1}{n_{miss}} \sum_{i=1}^{n_{miss}} \frac{|e_{miss}|}{|e_{miss}| + |e_{miss}^*|}$$
(8)

where e_{miss} and e_{miss}^* are the errors in two different methods. Based on this parameter, we compute the UMBRAE as:

$$UMBRAE = \frac{MBRAE}{1 - MBRAE} \tag{9}$$

When UMBRAE is equal to 1, the proposed method performs roughly as well as the reference method. When UMBRAE < 1, the assessed method performs roughly $(1 - \text{UMBRAE}) \times 100\%$ better than the reference method and if UMBRAE > 1, the method is roughly (UMBRAE – 1) × 100% worse than the reference. In this study, the i-ZOIB technique was the proposed method, and its performance was compared with those of the remaining imputation approaches. For instance, an i-ZOIB/LOCF UMBRAE value below 1 means that i-ZOIB performed better than the LOCF; if UMBRAE > 1, then i-ZOIB performed worse.

This was applied to each of the 250 runs. Thus, we computed the summary statistics (mean and standard deviations, as well as the median and the 25th and the 75th percentiles) in order to assess the imputation method performance.

3.4. Simulation Results

Simulation results for all the assessed scenarios, for the different missingness rates and mechanisms, are displayed in Tables A1–A6 (Appendix A). Bold numbers show the lowest MAE and RMSE values on each missingness rate and mechanism.

For each simulated scenario, 250 distinct samples with missing data were generated, with different missingness mechanisms and rates. Then, they were imputed. Finally, these new distributions were compared with those not containing imputed data, yielding the results depending on the assessed scenario.

Scenario 1: inverse-J-shaped distribution both at the baseline and at the follow-up

The MAE and RMSE results for the inverse-J-shaped imputed and not-imputed distributions are displayed in supplementary Tables A1 and A2. Boxplots of the distribution of the MAE accuracy measure are also depicted in Figure 3.

In general terms, the mean value of MAE was less than 0.07, both in the MCAR and MAR mechanisms and in whatever the percentage of data loss was. However, when the unobserved values followed an MNAR distribution, the SVR, 1-NN and LightGBoost methods showed mean MAE values greater than 0.10 (eg, MNAR 10% data loss: SVR: 0.09, LightGBoost: 0.106; MNAR 30%: 1-NN: 0.099, LightGBoost: 0.121). On the other hand, the ANN method showed the smallest mean MAE values of all the techniques compared in all the combinations of mechanisms and percentages of data losses. More precisely, when the missing values had a pattern of the MNAR type, the i-ZOIB method was also, together with the ANN, the one that presented the lowest error when imputing the missing values for whatever the missing rate was.

The SVR method presents a variability of MAE values higher than the rest of the methods in most of the situations studied, especially if the data imputation is carried out under the MAR or MNAR mechanism (Figure 3).

Regarding the RMSE findings, the same results are observed as in the MAE values. As the missing value mechanism depends more on the observed and unobserved data, the RMSE values increase: the MCAR case, the RMSE ranges between 0.067 (ANN, 10% loss) and 0.09 (1-NNI), while for that in the MNAR, the results are between 0.084 (ANN, 10% loss) and 0.17 (SVR, 30%) (Table A2).

In the second step, boxplots of UMBRAE results were constructed for the different fixed missingness rates and mechanisms (Figure 4). Based on the findings provided by the UMBRAE boxplots, all methods had similar performance to i-ZOIB. However, XGBOOST showed UMBRAE values less than 1 in the nine mechanism-percentage loss combinations studied in this data simulation. In situations where the loss of information was 30% (and any mechanism), the RF method showed UMBRAE values greater than 1.

Overall, its performance was comparable to that of the best method among them.

Scenario 2: inverse-J-shaped distribution at the baseline and inverse U-shaped distribution at the follow-up.

In the second evaluated setting, MAE and RMSE results for the imputed inverse Ushaped distribution and non-imputed distribution were compared. The data are displayed in Tables A3 and A4. MAE values in Scenario 1 under MCAR and at 10% of missingness rate

MAE values in Scenario 1 under MCAR and at 30% of missingness rate



MAE values in Scenario 1 under MCAR and at 20% of missingness rate

Figure 3. Boxplots of the MAE accuracy measure according to the missingness mechanisms, rates and imputation methods (from left to right: SVR, 1-NN, i-ZOIB, RF, ANN, XGBOOST, LightGBoost) in the Scenario 1.

The mean MAE values under the MCAR mechanism ranged between values of 0.054 (ANN, any percentage of loss) and 0.072 (XGBOOST, 10% of information loss). Similarly, the variability of MAE ranged from 0.003 (ANN, 10% loss) to 0.008 (1-NNI, 10% loss). Focusing on the results in the MAR mechanism and the missingness percentage of 20 –30%, it is observed that the SVR and ANN methods present the lowest MAE values, followed by those obtained by i-ZOIB (20% loss: SVR: 0.06; i-ZOIB: 0.01; ANN: 0.055; 30% loss: SVR: 0.06; i-ZOIB: 0.062; ANN: 0.057). Finally, and similarly to Scenario 1, the i-ZOIB and ANN methods showed the lowest values of the set of methods studied (Figure 5).



Figure 4. Boxplots of UMBRAE values according to the missingness mechanisms and rates in the Scenario 1.



Figure 5. Boxplots of the MAE accuracy measure according to the missingness mechanisms, rates, and imputation methods (from left to right: SVR, 1-NN, i-ZOIB, RF, ANN, XGBOOST, LightGBoost) in the Scenario 2.

Table A4 displays the results of the RMSE accuracy measures obtained in Scenario 2. Regarding the RMSE measure, its values were different according to the data loss mechanism: in both MCAR and MAR missingness mechanisms, all methods showed RMSE values lower than 0.10, while imputing the missing values with an MNAR distribution, these were higher than 0.10, except for ANN (in all percentages of unobserved values) and i-ZOIB (RMSE: 10% missingness rate: 0.085; 20% of missing data: 0.98). The methods that presented the lowest RMSE, by type of pattern, were the following: SVR and ANN (MCAR); SVR, RF and ANN (MAR); i-ZOIB and ANN (MNAR).

The boxplots of UMBRAE performance measurements for fixed missingness rates and mechanisms were constructed (Figure 6). Overall, the machine-learning and AI-based algorithms showed similar performance respect to the reference (i-ZOIB). Nevertheless, under the MNAR mechanism at 10% and 20% missingness rate, the RF, XGBoost and LightGBoost methods showed lower UMBRAE values than 1. On the other hand, at 30% of loss of information (in any mechanism), the RF method showed UMBRAE values higher than 1.





Figure 6. Boxplots of UMBRAE values according to the missingness mechanisms and rates in the Scenario 2.

• Scenario 3: J-shaped distribution both at the baseline and at follow-up.

Summary statistics of the MAE and RSME accuracy measures according to the missingness mechanisms and rates are displayed in Tables A5 and A6, respectively. Their corresponding boxplots are depicted in Figure 7.

In the current scenario, the mean MAE values ranged between 0.026 (LigthGBoost, MNAR, 10% losses) and 0.12 (SVR, MNAR, 30% missing values). Both in the MCAR and MAR mechanisms, i-ZOIB showed higher values than the rest of the techniques (MAE value: 0.06, versus 0.03–0.04 in the rest of the proposals). Furthermore, the variability of the MAE observed in each of the imputation methods was similar (Figure 7). On the other hand, in MNAR, the SVR method was the one who presented the worst MAE values (MAE: 30% losses: 0.12 vs. 0.02–0.03 the rest of the methods), also showing a higher variability than the rest of the techniques.

Boxplots show the distributions of UMBRAE values by the imputation method, mechanism, and missingness rate (Figure 8). In the MCAR mechanism, it is observed that the ML/AI techniques show UMBRAE values higher than 1, while in the MAR scenario, the 1-NN method is the only one that exceeds the threshold of 1. However, in the MNAR setting, both SVR, ANN and those based on boosting algorithms—XGBOOST, LightGBoost presented values below 1. Both 1-NN and RF continued to reflect values beyond 1.



Figure 7. Boxplots of the MAE accuracy measure according to the missingness mechanisms, rates and imputation methods (from left to right: SVR, 1-NN, i-ZOIB, RF, ANN, XGBOOST, LightGBoost) in the Scenario 3.



Figure 8. Boxplots of UMBRAE values according to the missingness mechanisms and rates in the Scenario 3.

4. Application Study

4.1. Description of the Data Set

The CARESS-CCR study started in 2010 and is still in progress. Patients with colorectal cancer (CRC) who had undergone the surgical procedure have been consecutively recruited and followed-up. The study has been approved by the research committee of the corresponding centres, and the patients have provided their informed consent to participate in the study.

The socio-demographic data and clinical variables were obtained. The patient-reported outcomes (PRO) were also included in the study. The participants could report their healthrelated quality of life (HRQoL) status by filling the appropriate questionnaires, starting at the baseline measurement point. A subset of patients with no missing values in the covariate related to the outcomes and the dropout process was selected. Moreover, to focus on the repeated study design to achieve our objectives, we selected the PRO measures to be examined at baseline and one year after the index surgical intervention.

We employed the k-NNI, SVR, and i-ZOIB methods to impute missing values in this subset of the outcome. We used the already-mentioned naïve techniques as control methods.

Our outcomes of interest and variables to be imputed were of two types. The first one was the total score in the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire, Core30 (QLQ-C30) [41,42]. This summary score (ranging from 0 to 100) covers all symptom (e.g., fatigue, pain) and function domains (e.g., emotional and social functioning) obtained for the cancer patients using the QLQ-C30. A single, higher-order HRQoL score has been proposed as a more meaningful and reliable measure in oncological research, where the higher the value, the better status of the patient. We also considered the anxiety level (HADS-A) reported by the patients using the HADS questionnaire [43,44]. It ranges from 0 to 21, where high scores reflect high anxiety levels and a reduction in well-being status. As covariates, we also selected the patient baseline HRQoL status and their scores in the health status.

The differences between the distributions obtained by the proposed imputation methods were evaluated.

4.2. Results

The study sample was composed of 968 patients. All participants reported their EORTC QLQ-C30 and HADS-A scores at the baseline point. Three hundred subjects (31%) did not fill the EORCT QLQ-C30 questionnaire a year after the surgical intervention. Similarly, 310 (32%) did not report their HAD-A levels during the follow-up. Figure 9 shows both the baseline and follow-up distributions of the HADS-A and EORTC QLQ-C30 scores. The former had a distribution (Scenario 3 of the simulation study), whereas the latter formed J-shaped curves at both time points (Scenario 1 in the simulation study).

To decide which imputation model to use, a logistic regression procedure was performed to assess the missingness mechanism related to the EORTC QLQ-C30 dropout during a year of follow-up. The results indicated that both the baseline EOTRC QLQ-C30 and HADS-A scores were associated with the dropout process. Likewise, both scores were related to the HADS-A follow-up dropout.

The distributions of the scores after applying different imputation methods are displayed in Figures 10 and 11. For both outcomes, imputing the missing values using the XGBOOST method led to biased distributions since the imputed values are outside the natural range of the imputed variable. The i-ZOIB method as well as the ANN, 1-NN, RF, and 1-NN imputation techniques provided similar curve shapes.



Figure 9. Distribution of the EORTC QLQ-C30 and HADS-A outcomes at the baseline and follow-up.



Figure 10. Distribution of the observed EORTC QLQ-C30 scores at follow-up (**green**) and after applying the imputation methods (**purple**).



Figure 11. Observed HADS- A score distribution at follow-up (**green**) and different imputed distributions after having applied the imputation methods (**purple**).

5. Discussion

Here, we evaluated the performance of several statistical and machine-learning approaches for handling missing data in bounded and non-bell-shaped repeated outcomes, a problem not often analysed so far. Our findings show that the i-ZOIB technique and ANN, SVR and RF machine-learning methods outperform the others in the accuracy assessments performed for different missingness mechanisms and rates in each of the evaluated scenarios.

Imputation using naïve methods such as the mean, median, or the LOCF technique is widely used. These naïve techniques ignore the correlation between the variables and underestimate the variability of the variable itself [45].

In the simulation study, we used the i-ZOIB method. This is a zero-one-inflated beta regression (based on Bayesian analysis) combined with the predictive mean matching methodology. The missing value is imputed using an observed value from a donor pool of k-candidate donors. i-ZOIB has several advantages over our other selected methods. This regression is appropriate for modelling outcomes with values within the [0,1] interval, as noted in the Methods section. This ensures that the predicted values remain within the same interval. It also introduces variability in the predicted values, as it is based on a Bayesian regression model. It is combined with the PMM method, which relaxes some of the assumptions of parametric imputation. Similar opinions have been presented in analogous studies [46]. Other statistical approaches have been considered to address this issue [27], but under another distinct setting. To our knowledge, it is the first approach for handling missing bounded outcomes in a repeated measurement setting.

In this work, we have compared the performance of machine learning and artificial intelligence techniques to address the treatment of missing values. In addition to the K-NN, the traditional algorithms ANN, SVR, RF, and those based on gradient descendent (XGBOOST and LightGBoost) have been applied. The results obtained in the simulation

study indicate that the mean RMSE values obtained by ANN in the different scenarios contemplated were lower than reported the rest of the techniques. This may be due to the fact that the SVR, RF, XGBOOST and LightGBoost methods are based on the feature selection, where they only consider the influence of each of the variables to establish the prediction. On the contrary, the ANN models evaluate the information considering the fusion of all the variables together. Furthermore, overestimation could be avoided (RFs are prone to provide overestimated models) by using the regularization method and the activation function during the training process. In our case, we have used the softplus function, a smoothed version of the well-known ReLu. We also conducted the imputation procedures using the K-NN approach, which in our case, 1-NN has been performed. This method, compared to other ML/AI methods, did not provide good results in MAE and RMSE assessments. As a single-imputation machine-learning technique, it takes into account relationships between variables, based on distance measurements.

In our results, both in the simulation and in the application studies, the XGBOOST method reflected imputed values beyond the range of values that the variable of interest is defined. This may be due to the design of the method itself. After the first iteration, each of the following trees is based on the error that occurred in the previous step/tree. For this reason, while the initial tree is restricted to the domain of the variable to be imputed, summing across gradient trees is not [31,47].

Another of the most important aspects is the type of pattern or distribution in which the missing values follow. The MCAR and MAR mechanisms either are patterns where the loss of information is established randomly (MCAR) or is conditioned by the observed values (MAR). No less important is the MNAR. In this case, the missing values not only depend on the observed information but also on the unobserved one. MNAR is non-ignorable, because to do so would lead to biased results. Therefore, knowing the mechanism is useful in identifying the most appropriate analysis. Our findings indicate that there is no single method in each of the data loss patterns and percentages, since, in each combination, we find, in addition to the ANN, other methods such as SVR or i-ZOIB that present similar mean MAE values. Similar results have been obtained. Similar results were concluded in different settings [13–23] when handling missing data using ML/AI techniques, depends on the scenario and type of the data structure to be imputed.

We used MAE, RMSE, and UMBRAE accuracy measures to assess the performance of the chosen imputation methods. Remarkably, UMBRAE is a good index to compare all the benchmarks, given a reference method. It combines the best features of various alternative measures into a single new one: it is a simple method, flexible and easy to use and understand and resistant to outliers. Moreover, the benchmark for calculating UMBRAE is selectable, and the ideal choice should be a forecasting method to be outperformed.

The results of this study can be only extrapolated to settings similar to those considered in our simulation, i.e., the three plausible real-life scenarios where the outcomes (bounded data) followed a heavy-tailed distribution.

6. Conclusions

We compared statistical and machine-learning approaches for handling missing data in skewed, bounded, and repeated outcomes, using plausible simulated real-life data. In the most frequent real-life settings (Scenario 1 and Scenario 3), the MAE and RMSE assessments of the i-ZOIB (statistical) as well as the ANN, SVR, and RF and (machinelearning) imputation methods resulted in low values. These methods can handle high missingness rates. However, their MAE and RMSE results could be improved; this might be achieved by optimising the parameters for each of these approaches.

In conclusion, our results can significantly contribute to improvements in the accuracy of analysis of bounded heavy-tailed and repeated outcomes. This would benefit many future research endeavours and real-life applications, such as modelling repeated HRQoL data or cost-effectiveness studies. **Author Contributions:** Conceptualization, U.A.-L. and C.E.B.; data curation, U.A.-L.; formal analysis, methodology and writing—original draft preparation, U.A.-L. and C.E.B.; writing—review and editing, U.A.-L. and C.E.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by a research grant from Instituto de Salud Carlos III (PI13/00013, PI18/01589) to U. Aguirre; Department of Health of the Basque Country (2010111098); KRONIKGUNE, Institute for Health Services Research (KRONIK 11/006); and the European Regional Development Fund. These institutions had no further role in study design; in the collection, analysis or interpretation of data; in the writing of the manuscript; or in the decision to submit the paper for publication.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the application study.

Data Availability Statement: The dataset used for the application study is available from the corresponding author on reasonable request.

Acknowledgments: The authors extend their appreciation to Jose María Quintana. The authors would like to thank the anonymous referees for their valuable comments and efforts which definitely help to improve the quality of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Supplementary Tables displaying MAE and RMSE values obtained on each of the studied scenario.

ANN

XGBOOST

LightGBoost

30

30

30

0.054

0.071

0.059

0.003

0.004

0.003

0.054

0.071

0.059

0.052

0.068

0.056

0.054

0.071

0.059

							Ū	0			*					
				MCAR					MAR					MNAR		
Imputation Method	%	Mean	SD	Median	P ₂₅	P ₇₅	Mean	SD	Median	P ₂₅	P ₇₅	Mean	SD	Median	P ₂₅	P ₇₅
SVR	10	0.055	0.006	0.055	0.052	0.055	0.062	0.007	0.061	0.056	0.062	0.099	0.020	0.097	0.083	0.099
1-NN	10	0.069	0.008	0.070	0.064	0.069	0.078	0.009	0.078	0.071	0.078	0.094	0.012	0.094	0.085	0.094
i-ZOIB	10	0.058	0.006	0.058	0.054	0.058	0.068	0.007	0.068	0.064	0.068	0.077	0.006	0.077	0.073	0.077
RF	10	0.057	0.006	0.057	0.054	0.057	0.061	0.007	0.061	0.057	0.061	0.093	0.013	0.093	0.084	0.093
ANN	10	0.054	0.005	0.054	0.051	0.054	0.058	0.006	0.058	0.054	0.058	0.068	0.010	0.066	0.060	0.068
XGBOOST	10	0.072	0.007	0.072	0.068	0.072	0.073	0.007	0.073	0.068	0.073	0.087	0.010	0.086	0.079	0.087
LightGBoost	10	0.058	0.006	0.058	0.054	0.058	0.066	0.008	0.065	0.060	0.066	0.106	0.010	0.106	0.099	0.106
SVR	20	0.056	0.004	0.055	0.053	0.056	0.060	0.006	0.060	0.056	0.060	0.110	0.018	0.108	0.097	0.110
1-NN	20	0.070	0.005	0.070	0.067	0.070	0.074	0.006	0.074	0.071	0.074	0.098	0.009	0.098	0.092	0.098
i-ZOIB	20	0.058	0.004	0.058	0.055	0.058	0.062	0.005	0.061	0.058	0.062	0.080	0.005	0.079	0.076	0.080
RF	20	0.058	0.004	0.058	0.055	0.058	0.061	0.005	0.060	0.058	0.061	0.096	0.011	0.096	0.089	0.096
ANN	20	0.054	0.003	0.054	0.052	0.054	0.055	0.005	0.054	0.051	0.055	0.070	0.026	0.068	0.062	0.070
XGBOOST	20	0.071	0.005	0.071	0.068	0.071	0.072	0.006	0.072	0.068	0.072	0.092	0.010	0.091	0.085	0.092
LightGBoost	20	0.059	0.004	0.058	0.056	0.059	0.064	0.005	0.063	0.060	0.064	0.111	0.010	0.111	0.105	0.111
SVR	30	0.056	0.004	0.055	0.053	0.056	0.060	0.006	0.060	0.056	0.060	0.120	0.021	0.117	0.105	0.120
1-NN	30	0.070	0.005	0.070	0.067	0.070	0.074	0.006	0.074	0.071	0.074	0.099	0.010	0.098	0.093	0.099
i-ZOIB	30	0.058	0.004	0.058	0.055	0.058	0.062	0.005	0.061	0.058	0.062	0.086	0.008	0.085	0.080	0.086
RF	30	0.058	0.003	0.058	0.055	0.058	0.061	0.003	0.061	0.059	0.061	0.099	0.012	0.098	0.090	0.099

Table A1. Summary statistics of the MAE accuracy measure according to the missingness mechanisms, rates and imputation methods in the Scenario 1.

%: Missingness rate; **SD**: Standard deviation; **P**₂₅: 25th percentile; **P**₇₅: 75th percentile. **SVR**: Support vector regression; **1-NN**: 1-Nearest neighbour; **i-ZOIB**: imputation via zero-one-inflated beta regression. RF: Random Forest; **ANN**: Artificial Neural Networks; **XGBOOST**: Extreme Gradient Boosting; **LightGBoost**: Light Gradient Boosting model.

0.003

0.003

0.004

0.057

0.072

0.065

0.056

0.069

0.062

0.057

0.072

0.065

0.070

0.092

0.121

0.009

0.010

0.012

0.068

0.092

0.120

0.063

0.085

0.112

0.070

0.092

0.121

0.057

0.072

0.065

		2			5		U	Ũ			1					
				MCAR					MAR					MNAR		
Imputation Method	%	Mean	SD	Median	P ₂₅	P ₇₅	Mean	SD	Median	P ₂₅	P ₇₅	Mean	SD	Median	P ₂₅	P ₇₅
SVR	10	0.068	0.007	0.068	0.064	0.068	0.08	0.011	0.079	0.073	0.08	0.134	0.031	0.13	0.11	0.134
1-NN	10	0.09	0.008	0.089	0.084	0.09	0.099	0.01	0.099	0.091	0.099	0.118	0.016	0.115	0.105	0.118
i-ZOIB	10	0.073	0.008	0.073	0.068	0.073	0.084	0.008	0.085	0.079	0.084	0.093	0.008	0.093	0.088	0.093
RF	10	0.072	0.007	0.072	0.067	0.072	0.077	0.008	0.076	0.071	0.077	0.117	0.016	0.117	0.106	0.117
ANN	10	0.067	0.006	0.066	0.063	0.067	0.072	0.007	0.072	0.068	0.072	0.084	0.013	0.082	0.075	0.084
XGBOOST	10	0.09	0.009	0.09	0.084	0.09	0.091	0.008	0.091	0.086	0.091	0.106	0.012	0.105	0.097	0.106
LightGBoost	10	0.072	0.007	0.071	0.067	0.072	0.083	0.01	0.083	0.077	0.083	0.133	0.012	0.133	0.125	0.133
SVR	20	0.069	0.005	0.069	0.065	0.069	0.083	0.013	0.082	0.074	0.083	0.155	0.027	0.156	0.135	0.155
1-NN	20	0.091	0.005	0.091	0.087	0.091	0.097	0.007	0.096	0.092	0.097	0.125	0.013	0.126	0.116	0.125
i-ZOIB	20	0.074	0.005	0.074	0.07	0.074	0.079	0.007	0.079	0.075	0.079	0.099	0.006	0.099	0.094	0.099
RF	20	0.072	0.004	0.072	0.07	0.072	0.078	0.007	0.078	0.074	0.078	0.124	0.014	0.124	0.115	0.124
ANN	20	0.067	0.004	0.067	0.064	0.067	0.069	0.007	0.068	0.065	0.069	0.088	0.028	0.085	0.078	0.088
XGBOOST	20	0.09	0.006	0.09	0.085	0.09	0.09	0.007	0.091	0.086	0.09	0.114	0.012	0.112	0.105	0.114
LightGBoost	20	0.073	0.005	0.073	0.07	0.073	0.084	0.008	0.084	0.078	0.084	0.142	0.012	0.142	0.135	0.142
SVR	30	0.069	0.005	0.069	0.065	0.069	0.083	0.013	0.082	0.074	0.083	0.171	0.028	0.169	0.152	0.171
1-NN	30	0.091	0.005	0.091	0.087	0.091	0.097	0.007	0.096	0.092	0.097	0.128	0.013	0.127	0.12	0.128
i-ZOIB	30	0.074	0.005	0.074	0.07	0.074	0.079	0.007	0.079	0.075	0.079	0.109	0.011	0.107	0.101	0.109
RF	30	0.072	0.004	0.072	0.07	0.072	0.077	0.005	0.077	0.074	0.077	0.13	0.016	0.128	0.119	0.13
ANN	30	0.067	0.003	0.067	0.064	0.067	0.071	0.003	0.071	0.069	0.071	0.088	0.013	0.085	0.079	0.088
XGBOOST	30	0.089	0.004	0.09	0.086	0.089	0.09	0.004	0.09	0.087	0.09	0.116	0.013	0.114	0.106	0.116
LightGBoost	30	0.074	0.004	0.073	0.071	0.074	0.085	0.005	0.085	0.081	0.085	0.158	0.015	0.158	0.148	0.158

Table A2. Summary statistics of the RMSE accuracy measure according to the missingness mechanisms, rates and imputation methods in the Scenario 1.

		•			-		0	0			-					
				MCAR					MAR					MNAR		
Imputation Method	%	Mean	SD	Median	P ₂₅	P ₇₅	Mean	SD	Median	P ₂₅	P ₇₅	Mean	SD	Median	P ₂₅	P ₇₅
SVR	10	0.056	0.005	0.056	0.053	0.056	0.056	0.005	0.056	0.052	0.056	0.073	0.016	0.069	0.063	0.073
1-NN	10	0.077	0.008	0.076	0.071	0.077	0.075	0.008	0.075	0.07	0.075	0.078	0.01	0.076	0.072	0.078
i-ZOIB	10	0.066	0.006	0.066	0.062	0.066	0.069	0.007	0.069	0.064	0.069	0.069	0.012	0.068	0.061	0.069
RF	10	0.058	0.006	0.058	0.055	0.058	0.058	0.006	0.058	0.054	0.058	0.077	0.011	0.075	0.07	0.077
ANN	10	0.058	0.005	0.058	0.054	0.058	0.058	0.005	0.058	0.055	0.058	0.066	0.008	0.065	0.061	0.066
XGBOOST	10	0.068	0.007	0.068	0.064	0.068	0.068	0.007	0.068	0.063	0.068	0.072	0.008	0.072	0.067	0.072
LightGBoost	10	0.059	0.006	0.059	0.056	0.059	0.06	0.006	0.06	0.056	0.06	0.085	0.011	0.084	0.079	0.085
SVR	20	0.057	0.003	0.057	0.054	0.057	0.057	0.004	0.056	0.054	0.057	0.083	0.018	0.077	0.068	0.083
1-NN	20	0.078	0.005	0.078	0.074	0.078	0.076	0.005	0.076	0.072	0.076	0.086	0.011	0.084	0.078	0.086
i-ZOIB	20	0.067	0.004	0.067	0.064	0.067	0.071	0.005	0.071	0.068	0.071	0.076	0.007	0.077	0.071	0.076
RF	20	0.059	0.004	0.058	0.056	0.059	0.059	0.004	0.059	0.057	0.059	0.082	0.009	0.081	0.075	0.082
ANN	20	0.058	0.004	0.058	0.056	0.058	0.058	0.004	0.058	0.056	0.058	0.067	0.006	0.065	0.062	0.067
XGBOOST	20	0.068	0.005	0.068	0.065	0.068	0.068	0.005	0.068	0.065	0.068	0.077	0.009	0.075	0.071	0.077
LightGBoost	20	0.06	0.004	0.06	0.057	0.06	0.062	0.004	0.062	0.059	0.062	0.087	0.008	0.087	0.082	0.087
SVR	30	0.057	0.003	0.057	0.054	0.057	0.057	0.004	0.056	0.054	0.057	0.087	0.017	0.085	0.072	0.087
1-NN	30	0.078	0.005	0.078	0.074	0.078	0.076	0.005	0.076	0.072	0.076	0.089	0.01	0.088	0.082	0.089
i-ZOIB	30	0.067	0.004	0.067	0.064	0.067	0.071	0.005	0.071	0.068	0.071	0.079	0.005	0.08	0.076	0.079
RF	30	0.059	0.003	0.059	0.057	0.059	0.06	0.003	0.06	0.058	0.06	0.083	0.008	0.082	0.077	0.083
ANN	30	0.058	0.003	0.058	0.056	0.058	0.058	0.003	0.058	0.057	0.058	0.067	0.006	0.065	0.063	0.067
XGBOOST	30	0.069	0.004	0.069	0.066	0.069	0.069	0.004	0.069	0.067	0.069	0.079	0.008	0.077	0.073	0.079
LightGBoost	30	0.06	0.003	0.06	0.058	0.06	0.064	0.004	0.064	0.062	0.064	0.088	0.006	0.088	0.084	0.088

Table A3. Summary statistics of the MAE accuracy measure according to the missingness mechanisms, rates and imputation methods in the Scenario 2.

| | | | | - |
 | 0 | 0
 | |
 | - | |
 | |
 | |
|----|------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
 | |
 | MAR |
 | | |
 | MNAR |
 | |
| % | Mean | SD | Median | P ₂₅ | P ₇₅
 | Mean | SD
 | Median | P ₂₅
 | P ₇₅ | Mean | SD
 | Median | P ₂₅
 | P ₇₅ |
| 10 | 0.069 | 0.006 | 0.069 | 0.065 | 0.069
 | 0.068 | 0.005
 | 0.068 | 0.064
 | 0.068 | 0.089 | 0.021
 | 0.084 | 0.077
 | 0.089 |
| 10 | 0.097 | 0.009 | 0.096 | 0.091 | 0.097
 | 0.095 | 0.009
 | 0.096 | 0.09
 | 0.095 | 0.1 | 0.01
 | 0.099 | 0.094
 | 0.1 |
| 10 | 0.081 | 0.007 | 0.081 | 0.076 | 0.081
 | 0.085 | 0.007
 | 0.085 | 0.08
 | 0.085 | 0.085 | 0.012
 | 0.085 | 0.077
 | 0.085 |
| 10 | 0.072 | 0.006 | 0.072 | 0.068 | 0.072
 | 0.072 | 0.006
 | 0.072 | 0.068
 | 0.072 | 0.091 | 0.011
 | 0.09 | 0.084
 | 0.091 |
| 10 | 0.07 | 0.005 | 0.069 | 0.067 | 0.07
 | 0.07 | 0.005
 | 0.07 | 0.066
 | 0.07 | 0.08 | 0.009
 | 0.078 | 0.074
 | 0.08 |
| 10 | 0.085 | 0.008 | 0.085 | 0.08 | 0.085
 | 0.084 | 0.008
 | 0.084 | 0.078
 | 0.084 | 0.089 | 0.009
 | 0.089 | 0.083
 | 0.089 |
| 10 | 0.073 | 0.006 | 0.073 | 0.069 | 0.073
 | 0.074 | 0.007
 | 0.074 | 0.069
 | 0.074 | 0.101 | 0.013
 | 0.1 | 0.093
 | 0.101 |
| 20 | 0.069 | 0.003 | 0.07 | 0.067 | 0.069
 | 0.069 | 0.004
 | 0.069 | 0.067
 | 0.069 | 0.11 | 0.031
 | 0.098 | 0.085
 | 0.11 |
| 20 | 0.097 | 0.006 | 0.097 | 0.093 | 0.097
 | 0.096 | 0.006
 | 0.096 | 0.092
 | 0.096 | 0.108 | 0.012
 | 0.106 | 0.1
 | 0.108 |
| 20 | 0.082 | 0.005 | 0.083 | 0.079 | 0.082
 | 0.087 | 0.006
 | 0.087 | 0.084
 | 0.087 | 0.094 | 0.008
 | 0.095 | 0.089
 | 0.094 |
| 20 | 0.073 | 0.004 | 0.073 | 0.07 | 0.073
 | 0.073 | 0.004
 | 0.073 | 0.07
 | 0.073 | 0.1 | 0.012
 | 0.099 | 0.091
 | 0.1 |
| 20 | 0.071 | 0.004 | 0.07 | 0.068 | 0.071
 | 0.07 | 0.004
 | 0.07 | 0.068
 | 0.07 | 0.081 | 0.008
 | 0.08 | 0.075
 | 0.081 |
| 20 | 0.086 | 0.005 | 0.085 | 0.082 | 0.086
 | 0.085 | 0.006
 | 0.085 | 0.082
 | 0.085 | 0.095 | 0.011
 | 0.093 | 0.088
 | 0.095 |
| 20 | 0.073 | 0.004 | 0.074 | 0.07 | 0.073
 | 0.076 | 0.005
 | 0.075 | 0.073
 | 0.076 | 0.11 | 0.01
 | 0.11 | 0.103
 | 0.11 |
| 30 | 0.069 | 0.003 | 0.07 | 0.067 | 0.069
 | 0.069 | 0.004
 | 0.069 | 0.067
 | 0.069 | 0.122 | 0.032
 | 0.117 | 0.091
 | 0.122 |
| 30 | 0.097 | 0.006 | 0.097 | 0.093 | 0.097
 | 0.096 | 0.006
 | 0.096 | 0.092
 | 0.096 | 0.113 | 0.013
 | 0.109 | 0.104
 | 0.113 |
| 30 | 0.082 | 0.005 | 0.083 | 0.079 | 0.082
 | 0.087 | 0.006
 | 0.087 | 0.084
 | 0.087 | 0.098 | 0.006
 | 0.099 | 0.095
 | 0.098 |
| 30 | 0.073 | 0.003 | 0.073 | 0.071 | 0.073
 | 0.074 | 0.003
 | 0.074 | 0.072
 | 0.074 | 0.105 | 0.012
 | 0.103 | 0.096
 | 0.105 |
| 30 | 0.07 | 0.003 | 0.07 | 0.069 | 0.07
 | 0.071 | 0.003
 | 0.07 | 0.069
 | 0.071 | 0.082 | 0.009
 | 0.08 | 0.076
 | 0.082 |
| 30 | 0.086 | 0.004 | 0.086 | 0.083 | 0.086
 | 0.087 | 0.004
 | 0.086 | 0.084
 | 0.087 | 0.098 | 0.011
 | 0.095 | 0.091
 | 0.098 |
| 30 | 0.074 | 0.003 | 0.074 | 0.072 | 0.074
 | 0.079 | 0.004
 | 0.079 | 0.076
 | 0.079 | 0.115 | 0.009
 | 0.115 | 0.109
 | 0.115 |
| | % 10 10 10 10 10 10 10 10 10 20 20 20 20 20 20 20 20 20 30 30 30 30 30 30 30 30 30 30 30 30 30 | % Mean 10 0.069 10 0.097 10 0.081 10 0.072 10 0.071 10 0.072 10 0.073 20 0.069 20 0.097 20 0.082 20 0.073 20 0.073 20 0.073 20 0.073 20 0.073 30 0.069 30 0.097 30 0.069 30 0.073 30 0.069 30 0.097 30 0.082 30 0.097 30 0.082 30 0.073 30 0.082 30 0.073 30 0.086 30 0.074 | % Mean SD 10 0.069 0.006 10 0.097 0.009 10 0.081 0.007 10 0.072 0.006 10 0.072 0.006 10 0.072 0.006 10 0.07 0.005 10 0.073 0.006 20 0.069 0.003 20 0.097 0.006 20 0.073 0.004 20 0.071 0.004 20 0.073 0.004 20 0.073 0.004 30 0.069 0.003 30 0.097 0.006 30 0.097 0.006 30 0.097 0.006 30 0.097 0.006 30 0.097 0.006 30 0.073 0.003 30 0.073 0.003 30 0.073 0.003 | MCAR % Mean SD Median 10 0.069 0.006 0.069 10 0.097 0.009 0.096 10 0.081 0.007 0.081 10 0.072 0.006 0.072 10 0.072 0.006 0.072 10 0.07 0.005 0.069 10 0.073 0.006 0.073 20 0.069 0.003 0.07 20 0.097 0.006 0.097 20 0.082 0.005 0.083 20 0.071 0.004 0.073 20 0.086 0.005 0.085 20 0.073 0.004 0.074 30 0.069 0.003 0.07 30 0.069 0.003 0.07 30 0.069 0.003 0.07 30 0.069 0.003 0.07 30 0.082 0.005 | MCAR % Mean SD Median P25 10 0.069 0.006 0.069 0.065 10 0.097 0.009 0.096 0.091 10 0.081 0.007 0.081 0.076 10 0.072 0.006 0.072 0.068 10 0.072 0.006 0.072 0.068 10 0.073 0.006 0.072 0.068 10 0.073 0.006 0.073 0.069 20 0.069 0.003 0.07 0.067 20 0.097 0.006 0.097 0.093 20 0.073 0.004 0.073 0.07 20 0.071 0.004 0.07 0.068 20 0.073 0.004 0.07 0.068 20 0.073 0.004 0.07 0.067 30 0.069 0.003 0.07 0.067 30 0.073 <td>MCAR % Mean SD Median P25 P75 10 0.069 0.006 0.069 0.065 0.069 10 0.097 0.009 0.096 0.091 0.097 10 0.081 0.007 0.081 0.076 0.081 10 0.072 0.006 0.072 0.068 0.072 10 0.07 0.005 0.069 0.067 0.07 10 0.073 0.006 0.073 0.069 0.073 10 0.073 0.006 0.073 0.069 0.073 10 0.073 0.006 0.073 0.069 0.073 20 0.069 0.003 0.07 0.067 0.069 20 0.097 0.006 0.097 0.093 0.097 20 0.082 0.005 0.083 0.07 0.073 20 0.071 0.004 0.07 0.068 0.071</td> <td>MCAR Mean SD Median P25 P75 Mean 10 0.069 0.006 0.069 0.065 0.069 0.068 10 0.097 0.009 0.096 0.091 0.097 0.095 10 0.081 0.007 0.081 0.076 0.081 0.085 10 0.072 0.006 0.072 0.068 0.072 0.072 10 0.07 0.005 0.069 0.067 0.07 0.072 10 0.07 0.005 0.069 0.067 0.07 0.072 10 0.073 0.006 0.073 0.069 0.073 0.074 10 0.073 0.006 0.073 0.069 0.067 0.074 20 0.069 0.003 0.07 0.067 0.069 0.069 20 0.097 0.006 0.097 0.093 0.097 0.096 20 0.073 0.004 0.073 <t< td=""><td>MCAR MCAR SD Median P25 P75 Mean SD 10 0.069 0.006 0.065 0.069 0.068 0.005 10 0.097 0.009 0.096 0.091 0.097 0.095 0.009 10 0.081 0.007 0.081 0.076 0.081 0.085 0.007 10 0.072 0.006 0.072 0.068 0.072 0.006 10 0.07 0.005 0.069 0.067 0.07 0.005 10 0.07 0.005 0.069 0.067 0.07 0.07 10 0.073 0.006 0.073 0.069 0.074 0.007 20 0.069 0.003 0.07 0.067 0.069 0.069 0.004 20 0.097 0.006 0.097 0.093 0.097 0.096 0.006 20 0.073 0.004 0.073 0.073 0.073 0.073 </td></t<><td>MCAR MAR % Mean SD Median P25 P75 Mean SD Median 10 0.069 0.006 0.065 0.069 0.008 0.005 0.068 10 0.097 0.009 0.096 0.091 0.097 0.095 0.009 0.096 10 0.081 0.007 0.081 0.076 0.081 0.085 0.007 0.085 10 0.072 0.006 0.072 0.068 0.072 0.072 0.006 0.072 10 0.07 0.005 0.069 0.067 0.07 0.005 0.071 10 0.073 0.006 0.073 0.069 0.073 0.074 0.007 0.074 20 0.069 0.003 0.07 0.067 0.069 0.069 0.097 0.096 0.006 0.097 20 0.069 0.003 0.07 0.067 0.069 0.069 0.073 0.073</td><td>MCAR MAR % Mean SD Median P25 P75 Mean SD Median P25 10 0.069 0.006 0.069 0.065 0.069 0.068 0.005 0.068 0.005 0.068 0.005 0.068 0.005 0.068 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.005 0.085 0.081 0.085 0.085 0.081 0.005 0.068 0.072 0.006 0.072 0.066 0.072 0.066 0.072 0.066 0.072 0.066 0.073 0.074 0.007 0.074 0.066 10 0.073 0.060 0.073 0.069 0.069 0.004 0.067</td><td>MCAR MAR % Mean SD Median P25 P75 Mean SD Median P25 P75 10 0.069 0.006 0.065 0.069 0.068 0.005 0.068 0.094 0.099 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.088 0.085 0.088 0.085 0.088 0.085 0.088 0.085 0.088 0.085 0.072 0.066 0.072 0.066 0.072 0.066 0.072 0.066 0.072 0.066 0.072 0.066 0.074 0.069 0.074 0.069 0.067 0.069 0.067 0.069 0.069 0.067 0.069 0.096<td>MCAR MAR % Mean SD Median P25 P75 Mean SD Median P25 P75 Mean SD Median P25 P75 Mean 10 0.069 0.006 0.069 0.065 0.069 0.068 0.005 0.068 0.064 0.068 0.089 10 0.097 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.085 0.088 0.085 0.085 0.081 0.085 0.081 0.085 0.081 0.085 0.084 0.085 0.081 0.085 0.072 0.066 0.072 0.068 0.072 0.069 0.074 0.007 0.074 0.069 0.074 0.010 10 0.085 0.008 0.085 0.089 0.069 0.074 0.010 0.073<!--</td--><td>MCAR MCAR MAR % Mean SD Median P25 P75 Mean SD Median P25 P75 Mean SD Median P25 P75 Mean SD 10 0.069 0.006 0.069 0.065 0.069 0.095 0.009 0.096 0.091 0.011 0.011 10 0.097 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.011 10 0.081 0.007 0.081 0.076 0.081 0.085 0.007 0.085 0.088 0.085 0.081 0.012 10 0.072 0.0068 0.072 0.0066 0.072 0.0066 0.072 0.008 0.081 0.085 0.081 0.085 0.084 0.078 0.084 0.089 0.091 10 0.073 0.069 0.073 0.074 0.007 0.074 0.069 0.074 0.101 <td< td=""><td>MCAR MCAR MAR MNAR % Mean SD Median P25 P75 Mean SD Median P25 P77 Mean SD Median P25 P77 Mean SD Median P25 P77 Mean SD Median 10 0.069 0.006 0.069 0.065 0.069 0.095 0.009 0.095 0.1 0.01 0.099 10 0.097 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.095 0.1 0.01 0.099 10 0.081 0.077 0.081 0.072 0.072 0.068 0.072 0.068 0.072 0.068 0.072 0.088 0.081 0.089 0.011 0.091 10 0.085 0.088 0.085 0.084 0.008 0.084 0.073 0.069 0.073 0.074 0.069 0.074 0.101 0.013 0.11</td><td>Image: Normal and the state of the</td></td<></td></td></td></td> | MCAR % Mean SD Median P25 P75 10 0.069 0.006 0.069 0.065 0.069 10 0.097 0.009 0.096 0.091 0.097 10 0.081 0.007 0.081 0.076 0.081 10 0.072 0.006 0.072 0.068 0.072 10 0.07 0.005 0.069 0.067 0.07 10 0.073 0.006 0.073 0.069 0.073 10 0.073 0.006 0.073 0.069 0.073 10 0.073 0.006 0.073 0.069 0.073 20 0.069 0.003 0.07 0.067 0.069 20 0.097 0.006 0.097 0.093 0.097 20 0.082 0.005 0.083 0.07 0.073 20 0.071 0.004 0.07 0.068 0.071 | MCAR Mean SD Median P25 P75 Mean 10 0.069 0.006 0.069 0.065 0.069 0.068 10 0.097 0.009 0.096 0.091 0.097 0.095 10 0.081 0.007 0.081 0.076 0.081 0.085 10 0.072 0.006 0.072 0.068 0.072 0.072 10 0.07 0.005 0.069 0.067 0.07 0.072 10 0.07 0.005 0.069 0.067 0.07 0.072 10 0.073 0.006 0.073 0.069 0.073 0.074 10 0.073 0.006 0.073 0.069 0.067 0.074 20 0.069 0.003 0.07 0.067 0.069 0.069 20 0.097 0.006 0.097 0.093 0.097 0.096 20 0.073 0.004 0.073 <t< td=""><td>MCAR MCAR SD Median P25 P75 Mean SD 10 0.069 0.006 0.065 0.069 0.068 0.005 10 0.097 0.009 0.096 0.091 0.097 0.095 0.009 10 0.081 0.007 0.081 0.076 0.081 0.085 0.007 10 0.072 0.006 0.072 0.068 0.072 0.006 10 0.07 0.005 0.069 0.067 0.07 0.005 10 0.07 0.005 0.069 0.067 0.07 0.07 10 0.073 0.006 0.073 0.069 0.074 0.007 20 0.069 0.003 0.07 0.067 0.069 0.069 0.004 20 0.097 0.006 0.097 0.093 0.097 0.096 0.006 20 0.073 0.004 0.073 0.073 0.073 0.073 </td></t<> <td>MCAR MAR % Mean SD Median P25 P75 Mean SD Median 10 0.069 0.006 0.065 0.069 0.008 0.005 0.068 10 0.097 0.009 0.096 0.091 0.097 0.095 0.009 0.096 10 0.081 0.007 0.081 0.076 0.081 0.085 0.007 0.085 10 0.072 0.006 0.072 0.068 0.072 0.072 0.006 0.072 10 0.07 0.005 0.069 0.067 0.07 0.005 0.071 10 0.073 0.006 0.073 0.069 0.073 0.074 0.007 0.074 20 0.069 0.003 0.07 0.067 0.069 0.069 0.097 0.096 0.006 0.097 20 0.069 0.003 0.07 0.067 0.069 0.069 0.073 0.073</td> <td>MCAR MAR % Mean SD Median P25 P75 Mean SD Median P25 10 0.069 0.006 0.069 0.065 0.069 0.068 0.005 0.068 0.005 0.068 0.005 0.068 0.005 0.068 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.005 0.085 0.081 0.085 0.085 0.081 0.005 0.068 0.072 0.006 0.072 0.066 0.072 0.066 0.072 0.066 0.072 0.066 0.073 0.074 0.007 0.074 0.066 10 0.073 0.060 0.073 0.069 0.069 0.004 0.067</td> <td>MCAR MAR % Mean SD Median P25 P75 Mean SD Median P25 P75 10 0.069 0.006 0.065 0.069 0.068 0.005 0.068 0.094 0.099 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.088 0.085 0.088 0.085 0.088 0.085 0.088 0.085 0.088 0.085 0.072 0.066 0.072 0.066 0.072 0.066 0.072 0.066 0.072 0.066 0.072 0.066 0.074 0.069 0.074 0.069 0.067 0.069 0.067 0.069 0.069 0.067 0.069 0.096<td>MCAR MAR % Mean SD Median P25 P75 Mean SD Median P25 P75 Mean SD Median P25 P75 Mean 10 0.069 0.006 0.069 0.065 0.069 0.068 0.005 0.068 0.064 0.068 0.089 10 0.097 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.085 0.088 0.085 0.085 0.081 0.085 0.081 0.085 0.081 0.085 0.084 0.085 0.081 0.085 0.072 0.066 0.072 0.068 0.072 0.069 0.074 0.007 0.074 0.069 0.074 0.010 10 0.085 0.008 0.085 0.089 0.069 0.074 0.010 0.073<!--</td--><td>MCAR MCAR MAR % Mean SD Median P25 P75 Mean SD Median P25 P75 Mean SD Median P25 P75 Mean SD 10 0.069 0.006 0.069 0.065 0.069 0.095 0.009 0.096 0.091 0.011 0.011 10 0.097 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.011 10 0.081 0.007 0.081 0.076 0.081 0.085 0.007 0.085 0.088 0.085 0.081 0.012 10 0.072 0.0068 0.072 0.0066 0.072 0.0066 0.072 0.008 0.081 0.085 0.081 0.085 0.084 0.078 0.084 0.089 0.091 10 0.073 0.069 0.073 0.074 0.007 0.074 0.069 0.074 0.101 <td< td=""><td>MCAR MCAR MAR MNAR % Mean SD Median P25 P75 Mean SD Median P25 P77 Mean SD Median P25 P77 Mean SD Median P25 P77 Mean SD Median 10 0.069 0.006 0.069 0.065 0.069 0.095 0.009 0.095 0.1 0.01 0.099 10 0.097 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.095 0.1 0.01 0.099 10 0.081 0.077 0.081 0.072 0.072 0.068 0.072 0.068 0.072 0.068 0.072 0.088 0.081 0.089 0.011 0.091 10 0.085 0.088 0.085 0.084 0.008 0.084 0.073 0.069 0.073 0.074 0.069 0.074 0.101 0.013 0.11</td><td>Image: Normal and the state of the</td></td<></td></td></td> | MCAR MCAR SD Median P25 P75 Mean SD 10 0.069 0.006 0.065 0.069 0.068 0.005 10 0.097 0.009 0.096 0.091 0.097 0.095 0.009 10 0.081 0.007 0.081 0.076 0.081 0.085 0.007 10 0.072 0.006 0.072 0.068 0.072 0.006 10 0.07 0.005 0.069 0.067 0.07 0.005 10 0.07 0.005 0.069 0.067 0.07 0.07 10 0.073 0.006 0.073 0.069 0.074 0.007 20 0.069 0.003 0.07 0.067 0.069 0.069 0.004 20 0.097 0.006 0.097 0.093 0.097 0.096 0.006 20 0.073 0.004 0.073 0.073 0.073 0.073 | MCAR MAR % Mean SD Median P25 P75 Mean SD Median 10 0.069 0.006 0.065 0.069 0.008 0.005 0.068 10 0.097 0.009 0.096 0.091 0.097 0.095 0.009 0.096 10 0.081 0.007 0.081 0.076 0.081 0.085 0.007 0.085 10 0.072 0.006 0.072 0.068 0.072 0.072 0.006 0.072 10 0.07 0.005 0.069 0.067 0.07 0.005 0.071 10 0.073 0.006 0.073 0.069 0.073 0.074 0.007 0.074 20 0.069 0.003 0.07 0.067 0.069 0.069 0.097 0.096 0.006 0.097 20 0.069 0.003 0.07 0.067 0.069 0.069 0.073 0.073 | MCAR MAR % Mean SD Median P25 P75 Mean SD Median P25 10 0.069 0.006 0.069 0.065 0.069 0.068 0.005 0.068 0.005 0.068 0.005 0.068 0.005 0.068 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.005 0.085 0.081 0.085 0.085 0.081 0.005 0.068 0.072 0.006 0.072 0.066 0.072 0.066 0.072 0.066 0.072 0.066 0.073 0.074 0.007 0.074 0.066 10 0.073 0.060 0.073 0.069 0.069 0.004 0.067 | MCAR MAR % Mean SD Median P25 P75 Mean SD Median P25 P75 10 0.069 0.006 0.065 0.069 0.068 0.005 0.068 0.094 0.099 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.088 0.085 0.088 0.085 0.088 0.085 0.088 0.085 0.088 0.085 0.072 0.066 0.072 0.066 0.072 0.066 0.072 0.066 0.072 0.066 0.072 0.066 0.074 0.069 0.074 0.069 0.067 0.069 0.067 0.069 0.069 0.067 0.069 0.096 <td>MCAR MAR % Mean SD Median P25 P75 Mean SD Median P25 P75 Mean SD Median P25 P75 Mean 10 0.069 0.006 0.069 0.065 0.069 0.068 0.005 0.068 0.064 0.068 0.089 10 0.097 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.085 0.088 0.085 0.085 0.081 0.085 0.081 0.085 0.081 0.085 0.084 0.085 0.081 0.085 0.072 0.066 0.072 0.068 0.072 0.069 0.074 0.007 0.074 0.069 0.074 0.010 10 0.085 0.008 0.085 0.089 0.069 0.074 0.010 0.073<!--</td--><td>MCAR MCAR MAR % Mean SD Median P25 P75 Mean SD Median P25 P75 Mean SD Median P25 P75 Mean SD 10 0.069 0.006 0.069 0.065 0.069 0.095 0.009 0.096 0.091 0.011 0.011 10 0.097 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.011 10 0.081 0.007 0.081 0.076 0.081 0.085 0.007 0.085 0.088 0.085 0.081 0.012 10 0.072 0.0068 0.072 0.0066 0.072 0.0066 0.072 0.008 0.081 0.085 0.081 0.085 0.084 0.078 0.084 0.089 0.091 10 0.073 0.069 0.073 0.074 0.007 0.074 0.069 0.074 0.101 <td< td=""><td>MCAR MCAR MAR MNAR % Mean SD Median P25 P75 Mean SD Median P25 P77 Mean SD Median P25 P77 Mean SD Median P25 P77 Mean SD Median 10 0.069 0.006 0.069 0.065 0.069 0.095 0.009 0.095 0.1 0.01 0.099 10 0.097 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.095 0.1 0.01 0.099 10 0.081 0.077 0.081 0.072 0.072 0.068 0.072 0.068 0.072 0.068 0.072 0.088 0.081 0.089 0.011 0.091 10 0.085 0.088 0.085 0.084 0.008 0.084 0.073 0.069 0.073 0.074 0.069 0.074 0.101 0.013 0.11</td><td>Image: Normal and the state of the</td></td<></td></td> | MCAR MAR % Mean SD Median P25 P75 Mean SD Median P25 P75 Mean SD Median P25 P75 Mean 10 0.069 0.006 0.069 0.065 0.069 0.068 0.005 0.068 0.064 0.068 0.089 10 0.097 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.097 0.085 0.088 0.085 0.085 0.081 0.085 0.081 0.085 0.081 0.085 0.084 0.085 0.081 0.085 0.072 0.066 0.072 0.068 0.072 0.069 0.074 0.007 0.074 0.069 0.074 0.010 10 0.085 0.008 0.085 0.089 0.069 0.074 0.010 0.073 </td <td>MCAR MCAR MAR % Mean SD Median P25 P75 Mean SD Median P25 P75 Mean SD Median P25 P75 Mean SD 10 0.069 0.006 0.069 0.065 0.069 0.095 0.009 0.096 0.091 0.011 0.011 10 0.097 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.011 10 0.081 0.007 0.081 0.076 0.081 0.085 0.007 0.085 0.088 0.085 0.081 0.012 10 0.072 0.0068 0.072 0.0066 0.072 0.0066 0.072 0.008 0.081 0.085 0.081 0.085 0.084 0.078 0.084 0.089 0.091 10 0.073 0.069 0.073 0.074 0.007 0.074 0.069 0.074 0.101 <td< td=""><td>MCAR MCAR MAR MNAR % Mean SD Median P25 P75 Mean SD Median P25 P77 Mean SD Median P25 P77 Mean SD Median P25 P77 Mean SD Median 10 0.069 0.006 0.069 0.065 0.069 0.095 0.009 0.095 0.1 0.01 0.099 10 0.097 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.095 0.1 0.01 0.099 10 0.081 0.077 0.081 0.072 0.072 0.068 0.072 0.068 0.072 0.068 0.072 0.088 0.081 0.089 0.011 0.091 10 0.085 0.088 0.085 0.084 0.008 0.084 0.073 0.069 0.073 0.074 0.069 0.074 0.101 0.013 0.11</td><td>Image: Normal and the state of the</td></td<></td> | MCAR MCAR MAR % Mean SD Median P25 P75 Mean SD Median P25 P75 Mean SD Median P25 P75 Mean SD 10 0.069 0.006 0.069 0.065 0.069 0.095 0.009 0.096 0.091 0.011 0.011 10 0.097 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.091 0.011 10 0.081 0.007 0.081 0.076 0.081 0.085 0.007 0.085 0.088 0.085 0.081 0.012 10 0.072 0.0068 0.072 0.0066 0.072 0.0066 0.072 0.008 0.081 0.085 0.081 0.085 0.084 0.078 0.084 0.089 0.091 10 0.073 0.069 0.073 0.074 0.007 0.074 0.069 0.074 0.101 <td< td=""><td>MCAR MCAR MAR MNAR % Mean SD Median P25 P75 Mean SD Median P25 P77 Mean SD Median P25 P77 Mean SD Median P25 P77 Mean SD Median 10 0.069 0.006 0.069 0.065 0.069 0.095 0.009 0.095 0.1 0.01 0.099 10 0.097 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.095 0.1 0.01 0.099 10 0.081 0.077 0.081 0.072 0.072 0.068 0.072 0.068 0.072 0.068 0.072 0.088 0.081 0.089 0.011 0.091 10 0.085 0.088 0.085 0.084 0.008 0.084 0.073 0.069 0.073 0.074 0.069 0.074 0.101 0.013 0.11</td><td>Image: Normal and the state of the</td></td<> | MCAR MCAR MAR MNAR % Mean SD Median P25 P75 Mean SD Median P25 P77 Mean SD Median P25 P77 Mean SD Median P25 P77 Mean SD Median 10 0.069 0.006 0.069 0.065 0.069 0.095 0.009 0.095 0.1 0.01 0.099 10 0.097 0.009 0.096 0.091 0.097 0.095 0.009 0.096 0.095 0.1 0.01 0.099 10 0.081 0.077 0.081 0.072 0.072 0.068 0.072 0.068 0.072 0.068 0.072 0.088 0.081 0.089 0.011 0.091 10 0.085 0.088 0.085 0.084 0.008 0.084 0.073 0.069 0.073 0.074 0.069 0.074 0.101 0.013 0.11 | Image: Normal and the state of the |

Table A4. Summary statistics of the RMSE accuracy measure according to the missingness mechanisms, rates and imputation methods in the Scenario 2.

		-			-		-	-			-					
				MCAR					MAR					MNAR		
Imputation Method	%	Mean	SD	Median	P ₂₅	P ₇₅	Mean	SD	Median	P ₂₅	P ₇₅	Mean	SD	Median	P ₂₅	P ₇₅
SVR	10	0.036	0.004	0.035	0.032	0.036	0.031	0.004	0.031	0.028	0.031	0.091	0.029	0.093	0.077	0.091
1-NN	10	0.043	0.007	0.042	0.039	0.043	0.036	0.006	0.036	0.032	0.036	0.03	0.005	0.03	0.027	0.03
i-ZOIB	10	0.059	0.008	0.059	0.052	0.059	0.049	0.007	0.049	0.043	0.049	0.029	0.005	0.029	0.026	0.029
RF	10	0.036	0.005	0.036	0.033	0.036	0.031	0.004	0.031	0.028	0.031	0.027	0.004	0.026	0.024	0.027
ANN	10	0.037	0.004	0.036	0.034	0.037	0.033	0.004	0.033	0.03	0.033	0.031	0.004	0.031	0.028	0.031
XGBOOST	10	0.041	0.006	0.041	0.038	0.041	0.034	0.005	0.034	0.031	0.034	0.03	0.004	0.029	0.027	0.03
LightGBoost	10	0.035	0.005	0.035	0.032	0.035	0.03	0.004	0.03	0.027	0.03	0.026	0.004	0.026	0.023	0.026
SVR	20	0.035	0.003	0.035	0.033	0.035	0.031	0.003	0.031	0.03	0.031	0.11	0.018	0.108	0.097	0.11
1-NN	20	0.044	0.005	0.044	0.041	0.044	0.037	0.004	0.037	0.034	0.037	0.031	0.004	0.031	0.029	0.031
i-ZOIB	20	0.06	0.005	0.06	0.056	0.06	0.052	0.005	0.052	0.048	0.052	0.032	0.003	0.032	0.03	0.032
RF	20	0.036	0.003	0.036	0.034	0.036	0.031	0.003	0.031	0.029	0.031	0.028	0.002	0.027	0.026	0.028
ANN	20	0.036	0.003	0.036	0.035	0.036	0.034	0.003	0.033	0.032	0.034	0.032	0.003	0.031	0.03	0.032
XGBOOST	20	0.041	0.004	0.041	0.039	0.041	0.035	0.003	0.035	0.033	0.035	0.031	0.002	0.031	0.029	0.031
LightGBoost	20	0.035	0.003	0.035	0.033	0.035	0.03	0.003	0.03	0.029	0.03	0.027	0.002	0.027	0.025	0.027
SVR	30	0.035	0.003	0.035	0.033	0.035	0.031	0.003	0.031	0.03	0.031	0.12	0.021	0.117	0.105	0.12
1-NN	30	0.044	0.005	0.044	0.041	0.044	0.037	0.004	0.037	0.034	0.037	0.032	0.003	0.032	0.03	0.032
i-ZOIB	30	0.06	0.005	0.06	0.056	0.06	0.052	0.005	0.052	0.048	0.052	0.036	0.003	0.036	0.034	0.036
RF	30	0.036	0.002	0.036	0.034	0.036	0.032	0.002	0.032	0.03	0.032	0.028	0.002	0.028	0.027	0.028
ANN	30	0.037	0.002	0.037	0.035	0.037	0.034	0.002	0.034	0.032	0.034	0.032	0.002	0.032	0.031	0.032
XGBOOST	30	0.041	0.003	0.041	0.04	0.041	0.036	0.003	0.037	0.035	0.036	0.032	0.002	0.032	0.03	0.032
LightGBoost	30	0.036	0.003	0.036	0.034	0.036	0.031	0.002	0.031	0.03	0.031	0.028	0.002	0.028	0.026	0.028

Table A5. Summary statistics of the MAE accuracy measure according to the missingness mechanisms, rates and imputation methods in the Scenario 3.

		5			5		Ũ	Ũ			1					
				MCAR					MAR					MNAR		
Imputation Method	%	Mean	SD	Median	P ₂₅	P ₇₅	Mean	SD	Median	P ₂₅	P ₇₅	Mean	SD	Median	P ₂₅	P ₇₅
SVR	10	0.048	0.007	0.048	0.043	0.048	0.041	0.007	0.041	0.036	0.041	0.122	0.041	0.123	0.1	0.122
1-NN	10	0.064	0.009	0.064	0.057	0.064	0.055	0.009	0.054	0.049	0.055	0.048	0.008	0.046	0.042	0.048
i-ZOIB	10	0.08	0.009	0.08	0.074	0.08	0.071	0.009	0.071	0.064	0.071	0.045	0.007	0.045	0.04	0.045
RF	10	0.05	0.007	0.05	0.046	0.05	0.043	0.006	0.042	0.039	0.043	0.038	0.006	0.037	0.033	0.038
ANN	10	0.048	0.006	0.048	0.044	0.048	0.043	0.005	0.042	0.039	0.043	0.04	0.006	0.039	0.036	0.04
XGBOOST	10	0.058	0.008	0.058	0.053	0.058	0.049	0.007	0.049	0.044	0.049	0.042	0.006	0.041	0.037	0.042
LightGBoost	10	0.049	0.007	0.049	0.044	0.049	0.042	0.006	0.042	0.038	0.042	0.037	0.006	0.037	0.033	0.037
SVR	20	0.048	0.005	0.048	0.045	0.048	0.042	0.005	0.041	0.038	0.042	0.155	0.027	0.156	0.135	0.155
1-NN	20	0.065	0.006	0.065	0.061	0.065	0.057	0.006	0.058	0.052	0.057	0.05	0.006	0.049	0.045	0.05
i-ZOIB	20	0.082	0.006	0.082	0.078	0.082	0.075	0.007	0.076	0.071	0.075	0.05	0.005	0.05	0.047	0.05
RF	20	0.05	0.005	0.051	0.047	0.05	0.044	0.005	0.044	0.041	0.044	0.039	0.004	0.039	0.036	0.039
ANN	20	0.048	0.004	0.047	0.045	0.048	0.044	0.004	0.043	0.041	0.044	0.042	0.005	0.041	0.038	0.042
XGBOOST	20	0.058	0.006	0.058	0.055	0.058	0.051	0.005	0.051	0.047	0.051	0.044	0.004	0.043	0.041	0.044
LightGBoost	20	0.05	0.005	0.05	0.046	0.05	0.043	0.005	0.043	0.04	0.043	0.039	0.004	0.038	0.036	0.039
SVR	30	0.048	0.005	0.048	0.045	0.048	0.042	0.005	0.041	0.038	0.042	0.171	0.028	0.169	0.152	0.171
1-NN	30	0.065	0.006	0.065	0.061	0.065	0.057	0.006	0.058	0.052	0.057	0.051	0.005	0.05	0.047	0.051
i-ZOIB	30	0.082	0.006	0.082	0.078	0.082	0.075	0.007	0.076	0.071	0.075	0.056	0.005	0.056	0.052	0.056
RF	30	0.05	0.004	0.05	0.048	0.05	0.045	0.004	0.045	0.042	0.045	0.04	0.004	0.04	0.037	0.04
ANN	30	0.048	0.004	0.048	0.045	0.048	0.044	0.003	0.044	0.042	0.044	0.042	0.004	0.042	0.04	0.042
XGBOOST	30	0.059	0.004	0.059	0.056	0.059	0.052	0.004	0.052	0.049	0.052	0.045	0.004	0.045	0.042	0.045
LightGBoost	30	0.051	0.004	0.051	0.048	0.051	0.045	0.004	0.045	0.042	0.045	0.04	0.004	0.04	0.037	0.04
-																

Table A6. Summary statistics of the RMSE accuracy measure according to the missingness mechanisms, rates and imputation methods in the Scenario 3.

References

- Schmidt, C.O.; Struckmann, S.; Enzenbach, C.; Reineke, A.; Stausberg, J.; Damerow, S.; Huebner, M.; Schmidt, B.; Sauerbrei, W.; Richter, A. Facilitating harmonized data quality assessments. A data quality framework for observational health research data collections with software implementations in R. *BMC Med. Res. Methodol.* 2021, *21*, 63. [CrossRef]
- 2. Roderick, J.A.; Little, D.B.R. Statistical Analysis with Missing Data, 2nd ed.; John Wiley and Sons: Hoboken, NJ, USA, 2002.
- 3. Janssen, K.J.; Donders, A.R.; Harrell, F.E., Jr.; Vergouwe, Y.; Chen, Q.; Grobbee, D.E.; Moons, K.G. Missing covariate data in medical research: To impute is better than to ignore. *J. Clin. Epidemiol.* **2010**, *63*, 721–727. [CrossRef]
- 4. Ng, C.G.; Yusoff, M.S.B. Missing Values in Data Analysis: Ignore or Impute? Educ. Med. J. 2011, 3, e6-e11. [CrossRef]
- 5. Xie, H. Analyzing longitudinal clinical trial data with nonignorable missingness and unknown missingness reasons. *Comput. Stat. Data Anal.* **2012**, *56*, 1287–1300. [CrossRef]
- 6. Fairclough, D.L. Design and Analysis of Quality of Life Studies in Clinical Trials; Chapman & Hall/CRC: Boca Raton, FL, USA, 2010.
- 7. Buuren, S.v.; Francis, T. Flexible Imputation of Missing Data; Chapman & Hall/CRC: Boca Raton, FL, USA, 2019.
- 8. Panés, J.; Vermeire, S.; Dubinsky, M.C.; Loftus, E.V.; Lawendy, N.; Wang, W.; Salese, L.; Su, C.; Modesto, I.; Guo, X.; et al. Efficacy and Safety of Tofacitinib Re-treatment for Ulcerative Colitis After Treatment Interruption: Results from the OCTAVE Clinical Trials. *J. Crohn's* **2021**. [CrossRef]
- 9. Blazek, K.; van Zwieten, A.; Saglimbene, V.; Teixeira-Pinto, A. A practical guide to multiple imputation of missing data in nephrology. *Kidney Int.* 2021, 99, 68–74. [CrossRef] [PubMed]
- 10. Ghosh, A. Robust inference under the beta regression model with application to health care studies. *Stat. Methods Med. Res.* 2017, 28, 871–888. [CrossRef] [PubMed]
- 11. Liu, F.; Eugenio, E.C. A review and comparison of Bayesian and likelihood-based inferences in beta regression and zero-or-oneinflated beta regression. *Stat. Methods Med. Res.* **2016**, *27*, 1024–1044. [CrossRef] [PubMed]
- 12. Ferrari, S.; Cribari-Neto, F. Beta Regression for Modelling Rates and Proportions. J. Appl. Stat. 2004, 31, 799–815. [CrossRef]
- 13. Chen, J.-B.; Lee, W.-C.; Cheng, B.-C.; Moi, S.-H.; Yang, C.-H.; Lin, Y.-D. Impact of risk factors on functional status in maintenance hemodialysis patients. *Eur. J. Med. Res.* 2017, 22, 54. [CrossRef]
- 14. Nosratabadi, S.; Mosavi, A.; Duan, P.; Ghamisi, P.; Filip, F.; Band, S.; Reuter, U.; Gama, J.; Gandomi, A. Data Science in Economics: Comprehensive Review of Advanced Machine Learning and Deep Learning Methods. *Mathematics* **2020**, *8*, 1799. [CrossRef]
- 15. Soleymani, F.; Masnavi, H.; Shateyi, S. Classifying a Lending Portfolio of Loans with Dynamic Updates via a Machine Learning Technique. *Mathematics* **2020**, *9*, 17. [CrossRef]
- 16. Su, Y.-C.; Wu, C.-Y.; Yang, C.-H.; Li, B.-S.; Moi, S.-H.; Lin, Y.-D. Machine Learning Data Imputation and Prediction of Foraging Group Size in a Kleptoparasitic Spider. *Mathematics* **2021**, *9*, 415. [CrossRef]
- Lakshminarayan, K.; Harp, S.A.; Samad, T. Imputation of Missing Data in Industrial Databases. *Appl. Intell.* 1999, 11, 259–275. [CrossRef]
- 18. Gill, M.K.; Asefa, T.; Kaheil, Y.; McKee, M. Effect of missing data on performance of learning algorithms for hydrologic predictions: Implications to an imputation technique. *Water Resour. Res.* **2007**, *43*. [CrossRef]
- 19. Jerez, J.M.; Molina, I.; García-Laencina, P.J.; Alba, E.; Ribelles, N.; Martín, M.; Franco, L. Missing data imputation using statistical and machine learning methods in a real breast cancer problem. *Artif. Intell. Med.* **2010**, *50*, 105–115. [CrossRef]
- 20. Chakraborty, D.; Başağaoğlu, H.; Winterle, J. Interpretable vs. noninterpretable machine learning models for data-driven hydro-climatological process modeling. *Expert Syst. Appl.* **2021**, 170, 114498. [CrossRef]
- 21. Serrano-Hidalgo, C.; Guardiola-Albert, C.; Aguilera, H. Estimating extremely large amounts of missing precipitation data. *J. Hydroinform.* 2020, 22, 578–592. [CrossRef]
- KA, N.D.; Tahir, N.M.; Abd Latiff, Z.I.; Jusoh, M.H.; Akimasa, Y. Missing data imputation of MAGDAS-9's ground electromagnetism with supervised machine learning and conventional statistical analysis models. *Alex. Eng. J.* 2021, *61*, 937–947. [CrossRef]
- 23. Zhang, X.; Yan, C.; Gao, C.; Malin, B.A.; Chen, Y. Predicting Missing Values in Medical Data Via XGBoost Regression. *J. Healthc. Inform. Res.* **2020**, *4*, 383–394. [CrossRef]
- 24. Muñoz, J.F.; Rueda, M. New imputation methods for missing data using quantiles. *J. Comput. Appl. Math.* **2009**, 232, 305–317. [CrossRef]
- 25. Raja, P.S.; Thangavel, K. Missing value imputation using unsupervised machine learning techniques. *Soft Comput.* **2019**, *24*, 4361–4392. [CrossRef]
- 26. Lee, K.J.; Carlin, J.B. Multiple imputation in the presence of non-normal data. Stat. Med. 2017, 36, 606–617. [CrossRef]
- 27. Geraci, M.; McLain, A. Multiple Imputation for Bounded Variables. Psychometrika 2018, 83, 919–940. [CrossRef]
- 28. Hu, C.; Yeilding, N.; Davis, H.M.; Zhou, H. Bounded outcome score modeling: Application to treating psoriasis with ustekinumab. *J. Pharmacokinet. Pharmacodyn.* **2011**, *38*, 497–517. [CrossRef]
- 29. Baione, F.; Biancalana, D.; Angelis, P. An application of Zero-One Inflated Beta regression models for predicting health insurance reimbursement. *arXiv* 2020, arXiv:2011.09248.
- 30. Schenker, N.; Taylor, J.M.G. Partially parametric techniques for multiple imputation. *Comput. Stat. Data Anal.* **1996**, *22*, 425–446. [CrossRef]

- 31. Beretta, L.; Santaniello, A. Nearest neighbor imputation algorithms: A critical evaluation. *BMC Med. Inform. Decision Mak.* 2016, 16, 197–208. [CrossRef] [PubMed]
- 32. Vapnik, V.N. Statistical Learning Theory; John Wiley and Sons: Hoboken, NJ, USA, 1998.
- 33. Vapnik, V.N. The Nature of Statistical Learning Theory; John Wiley and Sons: Hoboken, NJ, USA, 2010.
- 34. Awad, M.; Khanna, R. Efficient Learning Machines Theories, Concepts, and Applications for Engineers and System Designers; Springer: Berlin/Heidelberg, Germany, 2015.
- 35. Breiman, L. Random Forests. Mach. Learn. 2001, 45, 5–32. [CrossRef]
- Nawar, S.; Mouazen, A. Comparison between Random Forests, Artificial Neural Networks and Gradient Boosted Machines Methods of On-Line Vis-NIR Spectroscopy Measurements of Soil Total Nitrogen and Total Carbon. Sensors 2017, 17, 2428. [CrossRef]
- Chen, T.; Guestrin, C. XGBoost. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13 August 2016; pp. 785–794.
- Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A highly efficient gradient boosting decision tree. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4 December 2017; pp. 3149–3157.
- 39. Hyndman, R.J.; Koehler, A.B. Another look at measures of forecast accuracy. Int. J. Forecast. 2006, 22, 679–688. [CrossRef]
- 40. Gao, Z.-K.; Chen, C.; Twycross, J.; Garibaldi, J.M. A new accuracy measure based on bounded relative error for time series forecasting. *PLoS ONE* **2017**, *12*, e0174202. [CrossRef]
- Husson, O.; Rooij, B.H.; Kieffer, J.; Oerlemans, S.; Mols, F.; Aaronson, N.K.; Graaf, W.T.A.; Poll-Franse, L.V. The EORTC QLQ-C30 Summary Score as Prognostic Factor for Survival of Patients with Cancer in the "Real-World": Results from the Population-Based PROFILES Registry. Oncologist 2019, 25, e722. [CrossRef] [PubMed]
- 42. Kasper, B. The EORTC QLQ-C30 Summary Score as a Prognostic Factor for Survival of Patients with Cancer: A Commentary. *The Oncol.* **2020**, *25*, e610. [CrossRef]
- 43. Zigmond, A.S.; Snaith, R.P. The Hospital Anxiety and Depression Scale. Acta Psychiatr. Scand. 1983, 67, 361–370. [CrossRef]
- 44. Herrero, M.J.; Blanch, J.; Peri, J.M.; De Pablo, J.; Pintor, L.; Bulbena, A. A validation study of the hospital anxiety and depression scale (HADS) in a Spanish population. *Gen. Hosp. Psychiatry* **2003**, *25*, 277–283. [CrossRef]
- 45. Buhi, E. Out of Sight, Not Out of Mind: Strategies for Handling Missing Data. *Am. J. Health Behav.* **2008**, *32*, 83–92. [CrossRef] [PubMed]
- 46. Kwon, T.Y.; Park, Y. A new multiple imputation method for bounded missing values. *Stat. Probab. Lett.* **2015**, 107, 204–209. [CrossRef]
- 47. Kim, T.; Ko, W.; Kim, J. Analysis and Impact Evaluation of Missing Data Imputation in Day-ahead PV Generation Forecasting. *Appl. Sci.* **2019**, *9*, 204. [CrossRef]