UDC 316.4

Kostenko, Y., & Gorbachyk, A. (2025). Missing Categorical Data in Sociological Surveys: An Experimental Evaluation of Imputation Techniques. *Sociological Studios*, 1(26), 85–109. https://doi.org/10.29038/2306-3971-2025-01-32-32

Missing Categorical Data in Sociological Surveys: An Experimental Evaluation of Imputation Techniques

Yaroslav Kostenko –
PhD student, Sociology faculty, Taras Shevchenko National University of Kyiv, Ukraine
E-mail: yarosl.kostenko@gmail.com
ORCID: 0009-0001-7878-5034

Andrii Gorbachyk –
Candidate of science in Mathematics, Associate Professor, Department of Methodology and Methods of Sociological Research, Sociology faculty, Taras Shevchenko National University of Kyiv, Ukraine
E-mail: a.gorbachyk@knu.ua
ORCID: 0000-0003-1944-435X

Ярослав Костенко –
PhD студент, факультет соціології, КНУ імені Тараса Шевченка

Андрій Горбачик –
кандидат фіз.-мат. наук, доцент, кафедра методології та методів соціологічних досліджень, факультет соціології, КНУ імені Тараса Шевченка

DOI: 10.29038/2306-3971-2025-01-32-32

Received: April 09, 2025
1st Revision: April 21, 2025
Accepted: June 18, 2025

Missing categorical data presents a persistent challenge to data quality in quantitative sociological research, where simpler approaches can lead to biased estimates and incorrect conclusions. This article provides an empirically grounded evaluation of multiple imputation (MI) strategies for categorical survey data, specifically focusing on the complex, multi-category nominal variable "party voted for" using European Social Survey data from Sweden and Norway. We developed a simulation framework, introducing missingness under Missing Completely at Random, Missing at Random, derived from patterns of item nonresponse on auxiliary variables, and Missing Not at Random: linked to the undisclosed party choice itself. We systematically compared the performance of six imputation methods (Multinomial Logistic Regression, Random Forest, CART, KNN, Hot Deck, and Mode) across four distinct predictor set sizes, evaluating them using Accuracy, Cohen's Kappa, and Macro F1-score with m=20 imputations. Results indicate that while imputing party choice is challenging, model-based MI techniques significantly outperform naive approaches. Multinomial Logistic Regression consistently emerged as the most robust and highest-performing method, often benefiting from larger predictor sets within the MI framework. K-Nearest Neighbors showed promise with smaller predictor sets, offering a computationally efficient alternative. The work emphasizes the importance of principled imputation and provides practical recommendations for sociologists regarding method selection, predictor set construction, and consideration of computational costs when addressing missing categorical data.

Key words: Data Quality, Missing Data, Data Imputation, Multiple Imputation, Logistic Regression, Clustering.

Костенко Ярослав, Горбачик Андрій. Пропущені категоріальні дані в соціологічних опитуваннях: експериментальна оцінка технік імпутації. Відсутність категоріальних даних залишається актуальною проблемою якості даних у кількісних соціологічних дослідженнях, де прості підходи можуть призвести до зміщень і хибних висновків. У цій статті представляємо емпірично обґрунтовану оцінку стратегій множинної імпутації (MI) для категоріальних даних опитувань, фокусуючись на мультикатегоріальну номінальну змінну «партія, за яку проголосував респондент», використовуючи дані European Social Survey зі Швеції та Норвегії. Ми розробили симуляційну схему, генеруючи пропущені дані за механізмами Missing Completely at Random, Missing at Random, на основі патернів відмов респондентів в інших змінних, та Missing Not at Random, що засновується на самому виборі партії. Систематично порівняно ефективність шести методів імпутації (Multinomial Logistic Regression, Random Forest, CART, KNN, Hot Deck і Mode) за чотирьох різних за обсягом наборів предикторів, оцінюючи їх за Accuracy, Cohen's Kappa та Macro F1-score при m=20 імпутаціях. Результати свідчать, що хоча імпутація партійного вибору є нетривіальною проблемою, MI з використанням предиктивних моделей суттєво перевершує більш прості підходи. Multinomial Logistic Regression послідовно показує найстійкіші й найвищі результати, часто виграючи від більших наборів предикторів у рамках MI. K-Nearest Neighbors демонструє перспективність за менших наборів предикторів, пропонуючи обчислювально ефективну альтернативу. У роботі підкреслюється важливість принципового підходу до імпутації та надаються

практичні рекомендації соціологам щодо вибору методу, побудови набору предикторів і врахування обчислювальних витрат під час роботи з пропущеними категоріальними даними.

Ключові слова: якість даних, пропущені дані, імпутація даних, множинна імпутація, логістична регресія, кластеризація.

INTRODUCTION

Missing data is a common challenge in quantitative social science research. It is often ignored or handled through simplistic approaches such as listwise deletion, which can lead to significant problems, including reduced sample size, biased estimates, and even false conclusions, especially when data are not Missing Completely at Random (MCAR). Under MCAR, the probability of a value being missing is entirely independent of both observed and unobserved data. For instance, a survey respondent might accidentally skip a question, or a data entry error might randomly delete a value, with no underlying systematic reason. However, in many sociological surveys, especially those involving sensitive topics, data are rarely MCAR. Instead, the missingness often follows an underlying pattern related to other information. For example, in surveys collecting income data, respondents in certain occupations (an observed variable) might be less likely to report their income, exemplifying a Missing at Random (MAR) pattern. If, additionally, individuals with very high or very low incomes (the unobserved income values themselves) are less likely to disclose this information, this illustrates a Missing Not at Random (MNAR) mechanism. In such MAR and MNAR scenarios, where both the proportion and the mechanism of missing data are non-trivial, simplistic approaches to missing data can lead to particularly misleading results.

While missing data is a widely recognized issue across disciplines such as medicine and economics, it presents unique challenges in sociology. Sociological data often rely on non-metric scales, such as ordinal Likert items or nominal classifications, which restrict the range of applicable imputation methods. Many standard techniques in other fields assume metric scales, for instance, continuous income values or physiological measurements. Such approaches are less fitting for common sociological variables. Although ordinal variables are sometimes treated as quasimetric, which would allow the usage of methods used for metric scales, categorical data require entirely different methodological approaches.

This article aims to systematically evaluate multiple imputation methods for categorical data and provide evidence-based recommendations tailored to social scientists. Several studies have compared imputation strategies for categorical variables, often focusing on dichotomous outcomes (Dong et al., 2021; Ge et al., 2023; Lang, & Wu, 2017). Our work extends this by using realistic, multi-category variables and by designing a simulation framework that reflects the complexities of actual survey data. This approach is closer to the real case scenarios in the quantitative sociological researches.

In order to achieve this, we begin with an ideal dataset drawn from the European Social Survey (ESS), Wave 11, picking a subset with no missing values on our target variable: “party voted for in the most recent national election”. We then introduce missingness to this specific variable in a controlled and theoretically grounded way:

1. Missing Completely at Random (MCAR): A portion of the “party voted for” responses are deleted purely at random. This simulates missingness due to entirely stochastic processes, such as accidental data entry skips or errors unrelated to any respondent characteristics or their actual vote.
2. Missing at Random (MAR): The likelihood of “party voted for” being missing is made dependent on other observed variables. We’ve clustered respondents based on their patterns of nonresponse to a set of auxiliary survey questions (e.g., on political and social views). Respondents in clusters with higher item nonresponse rates were assigned a higher probability of having their party choice deleted.
3. Missing Not at Random (MNAR): The probability of “party voted for” being missing is directly tied to the value of the party choice itself. For instance, responses indicating a vote for parties pre-defined as more controversial or less socially acceptable were assigned a higher probability of deletion, simulating respondents’ reluctance to disclose such choices.

This yields two evaluation scenarios: one with only MCAR/MAR patterns, and another incorporating a more difficult MNAR component, allowing us to test the robustness of each method under increasing complexity of the missingness mechanism.

We assess a range of commonly used imputation techniques for categorical data using three performance metrics: accuracy, F1 Macro, and Cohen’s Kappa.

- Accuracy provides a simple measure of the overall proportion of imputed party choices that exactly match the respondents' true, original party preferences. While intuitive and straightforward, accuracy can be misleading in multi-category scenarios, especially if some parties (categories) are much more prevalent than others.
- The Macro F1-score offers a more balanced perspective by calculating the F1-score (the harmonic mean of precision and recall) for each political party independently and then averaging these scores. This gives equal weight to the imputation performance for both large and small parties, making it sensitive to how well the method imputes less frequent choices.
- Cohen's Kappa is particularly important for this evaluation. It measures the agreement between the imputed and true party choices while explicitly correcting for the amount of agreement that would be expected by chance alone. In survey data with multiple nominal categories like party preference, where a naive imputation could achieve some accuracy purely by chance (e.g., by frequently guessing the largest party, especially if there's a single party with high support), Kappa provides a more conservative and reliable estimate of the imputation method's true discriminatory power.

Higher values across these metrics, especially for Cohen's Kappa, indicate a greater ability of the imputation method to correctly recover the specific party affiliations of individual respondents beyond what random chance would predict. Therefore, they reflect the method's success in accurately reconstructing the missing individual-level data points.

Our findings contribute both practical and theoretical insights. We provide concrete recommendations for handling missing categorical data in survey research, which is an often overlooked, but highly prevalent issue in quantitative sociological research.

1. CHOICE OF DATASET AND QUESTIONS OF INTEREST

We selected Wave 11 of the European Social Survey (ESS) as our primary data source, focusing on two countries: Sweden and Norway. Our target variable was "party voted for in the most recent national election" (later referred to as "party choice"), a question respondents may be hesitant to answer, particularly when their choice involves parties perceived as socially undesirable (ESS ERIC, 2024).

Importantly, both Sweden and Norway exhibit high response rates on this "party choice" item within the ESS. This characteristic was crucial, as it allowed us to construct a robust "ideal dataset" to serve as our baseline. This ideal dataset was formed by selecting only those respondents who indicated they had voted and also subsequently provided a valid answer to the 'party choice' question. By starting with this complete data from actual voters who disclosed their preferences, we can treat it as highly representative of this specific population group within these countries. This foundation is vital for accurately evaluating the performance of imputation methods, as the original (complete) data acts as the source of truth against which imputed values are compared. In contrast, beginning with data from countries with high initial nonresponse on the target variable would weaken the claim that our complete baseline truly reflects the population of interest.

Moreover, both Sweden and Norway feature diverse and fragmented political landscapes with a wide range of parties represented. This multi-party context makes them particularly appropriate for evaluating imputation techniques on multi-category nominal variables, as the inherent response variability poses a greater challenge for imputation models. The combination of high-quality baseline data (due to high initial response rates among voters who disclosed their choice) and complex multi-party systems enhances the external validity of our evaluation, making our findings more applicable to other sociological contexts where categorical data are both sensitive and diverse.

2. GENERATING THE MISSING DATA

The next key step involves generating missing data in a way that reflects realistic patterns of nonresponse. According to Rubin's (1976) widely used framework, known as Rubin's Classification System, missing data can be categorized into three types:

1. Missing Completely at Random (MCAR) refers to situations where the probability of a data point being missing is entirely independent of both observed and unobserved data. For example, when data is missing due to an encoding error, with no relation to any characteristic of the respondent or the value itself.

Table 1

Distribution of Valid Responses for Sweden and Norway

Country	**Sweden**			**Norway**		
Valid Percentage	87,6			77,6		
	Party	**N**	**%**	**Party**	**N**	**%**
	Centern	103	9,6	Rødt	52	5,0
	Kristdemokraterna	61	5,7	Sosialistisk Venstreparti	93	9,0
	Liberalerna	48	4,5	Arbeiderpartiet	284	27,4
	Miljöpartiet de gröna	67	6,2	Venstre	47	4,5
	Moderata samlingspartiet	210	19,5	Kristelig Folkeparti	32	3,1
	Socialdemokraterna	381	35,3	Senterpartiet	133	12,8
	Sverigedemokraterna	86	8,0	Høyre	248	23,9
	Vänsterpartiet	108	10,0	Fremskrittspartiet	76	7,3
	Other	14	1,3	Miljøpartiet De Grønne	43	4,2

Source: *compiled by the authors.*

2. Missing at Random (MAR) occurs when missingness is systematically related to observed variables, but not to the values of the missing data themselves. For example, when respondent omits a question related to discrimination due to their gender. This probability of omission is tied to the other observed variables, in this case–gender, for example, with women less likely to report discrimination.

3. Missing Not at Random (MNAR) describes cases where the probability of missingness is related to the unobserved value itself. For example, when a respondent omits a politically sensitive answer precisely because of having a socially condemnable answer for it. This means that the probability is tied to the answer itself: socially condemnable answers will have higher chance of being missing (Rubin, 1976).

However, as Graham (2009) points out, it is a misconception to treat MCAR, MAR, and MNAR as mutually exclusive or clear categories. In practice, datasets rarely conform perfectly to any one type, and real-world missingness tends to fall somewhere along a continuum, often on the spectrum between MAR and MNAR characteristics (Graham, 2009). Purely MCAR data is rare, and the assumptions required for strictly MAR mechanisms are often unrealistic. As such, it is more accurate to acknowledge that most missingness exhibits at least some degree of MNAR behavior.

In this study, we design our missing data generation mechanisms to reflect this complexity. We construct two test datasets:

1. The first includes only MCAR and MAR components, representing a less severe scenario of missing data, where MNAR component is negligible. An example could involve a non-sensitive survey item (e.g., preferred leisure activities) where any missing responses are presumed to occur randomly (MCAR) or are correlated with other observed respondent characteristics like age or education (MAR), rather than being due to reluctance to reveal the specific leisure activity itself.

2. The second introduces MNAR elements as well, simulating a more difficult and realistic missing data problem.

An important consideration in designing these mechanisms is that they must be non-trivial and opaque–that is, the pattern of missingness should not be easily learnable or "decoded" by imputation models. This ensures that imputation methods are genuinely evaluated based on their ability to predict the missing values based on patterns in the data, rather than exploiting clues from how missingness was introduced.

For this simulation of a realistic missing data scenario, we set the total proportion of missing values in our outcome variable to 20 %. This level of missingness is often identified in the literature as a point where

simpler approaches begin to fail and more robust, advanced imputation methods are required (Lee & Huber, 2021).

MCAR component is inherently random, and is generated by deleting a fraction of values randomly. For this case, we've settled on 20 % of missingness being MCAR, while the remaining 80 % would be either MAR or MNAR.

To construct the MAR component of missingness, we aimed to make the probability of missingness depend on observable response behavior. For this case, we've clustered the respondents according to their response patterns for a wide set of questions that had non-responses, such as political and cultural views (full list of questions available in the appendix A). These were used to encode respondents into a binary matrix (1 = answer, 0 = no answer), which formed the basis for clustering.

We applied three distinct clustering algorithms–Ward's method, K-Modes, and Gaussian Mixture Models (GMM)–to this binary matrix, exploring a range of cluster solutions (from 3 to 10 clusters) for each algorithm. To generate richer and more nuanced respondent typologies than a single algorithm might provide, while avoiding the excessive fragmentation of combining all three, we created "superclusters" by forming pairwise intersections of the cluster assignments from any two of these three algorithms. For instance, if a respondent was assigned to cluster "A" by one algorithm (e.g., K-Modes with 8 clusters) and cluster "B" by a second algorithm (e.g., GMM with 3 clusters), they were placed into the supercluster "A_B" (e.g. K-Modes8_GMM3). This process was repeated for all pairs of algorithms (Ward's & K-Modes, Ward's & GMM, K-Modes & GMM) across their various tested cluster numbers, yielding a large candidate pool of supercluster solutions. We then evaluated each supercluster combination using three criteria:

1. Normalized Shannon entropy of cluster sizes. This metric assessed the balance of the resulting respondent groupings. Higher entropy indicates more evenly sized superclusters, which is desirable for ensuring that our MAR mechanism is not disproportionately driven by a few very large or very small groups of respondents. A balanced solution suggests a more stable and generalizable typology of response behavior.

2. Normalized Weighted Standard Deviation (WSD) of supercluster-specific refusal rates. To ensure our superclusters captured genuine differences in nonresponse patterns, we calculated the average item nonresponse rate (based on the auxiliary questions) within each supercluster. The WSD of these rates across superclusters measures their differentiation. A higher WSD signifies that the superclusters effectively distinguish between groups of respondents with substantially different underlying tendencies to omit answers, which is important for creating a meaningful MAR mechanism linked to observed response patterns.

3. Fraction of clusters of insignificant size. This criterion penalized solutions that produced excessive fragmentation or noise. A supercluster was defined as 'tiny' if its size was less than 10 % of the average expected size had all superclusters in that particular solution been perfectly balanced (i.e., size < 0,1* (total sample size / number of superclusters)). A high fraction of such tiny clusters might reflect overfitting by the clustering algorithms or unstable groupings. We aimed to minimize this fraction to focus on more substantial patterns of nonresponse behavior.

For each country, we selected the supercluster solution that demonstrated the most favorable balance across our three evaluation criteria, aiming for a solution that was well-differentiated by nonresponse behavior (high WSD), structurally balanced (high entropy), and minimally fragmented (low tiny-cluster fraction). While no single solution typically maximizes all criteria simultaneously, we prioritized solutions that showed strong differentiation in refusal rates (high WSD) while maintaining reasonable balance and avoiding excessive numbers of tiny clusters:

1. In the case of Sweden, the optimal pairing was K-Modes (8 clusters) with GMM (3 clusters), achieving a normalized entropy of 0,66, weighted standard deviation (WSD) of 0,21, and a tiny-cluster fraction of 0,07.

2. For Norway, the best combination was K-Modes (6 clusters) and GMM (3 clusters), with corresponding values of 0,58, 0,23, and 0,08, respectively.

These chosen solutions represented the best compromise for creating meaningful, behaviorally distinct groups for our MAR simulation.

Each respondent was then assigned a deletion weight based on the average nonresponse rate within their supercluster. These weights were used to simulate the MAR aspect of missingness, where respondents

from groups with higher prior nonresponse patterns were more likely to have their outcome value (party choice) removed. This ensures that the MAR component of missingness is realistic, informed by known nonresponse behavior, and also not easily learnable by simple imputation models.

For example, if one respondent had a deletion weight of 0,1 and another 0,2, the second would be twice as likely to have their response removed.

To simulate the MNAR (Missing Not At Random) component–where the probability of missingness depends directly on the value of the variable itself–we introduced the assumption that respondents may be less likely to disclose their party preference if they support a party considered less socially acceptable or more politically controversial. We grouped parties in each country into three categories based on perceived social acceptability:

1. Mainstream parties, assumed to have no added likelihood of being omitted.
2. Somewhat mainstream parties, with slightly increased risk of nonresponse. This also included the "other party" option.
3. Controversial or fringe parties, with the highest assumed likelihood of nonresponse.

The grouping draws on established party-positioning research, most notably the 2019 Chapel Hill Expert Survey (Bakker et al., 2020) and specialist party profiles (Bjerkem, 2016; Bulent, 2020; Center for Strategic & International Studies, 2021; Jupskås, & Langsæther, 2023). Each category was assigned a numeric weight that adjusted the probability of deletion. These party-specific weights were then used to adjust the MAR-derived deletion probability for each respondent. Specifically, the MAR-based probability was multiplied by the assigned party weight, effectively amplifying the likelihood of deletion for respondents who chose parties in the higher-weighted categories (i.e., non-mainstream parties) relative to those who chose mainstream parties (weight 1,0).

In the Swedish dataset:

1. Mainstream parties such as Centern, Kristdemokraterna, Liberalerna, Moderaterna, Socialdemokraterna were assigned a weight of 1.0.
2. Somewhat mainstream or smaller parties (Miljöpartiet, Other) received a weight of 1,5.
3. Parties regularly labelled populist-radical right or radical left (Sverigedemokraterna, Vänsterpartiet) were given the weight of 2,0 (Bulent, 2020)

In the Norwegian dataset:

1. Established centre-left/centre-right parties such as Arbeiderpartiet, Venstre, Kristelig Folkeparti, Senterpartiet, and Høyre were assigned a weight of 1,0.
2. Integrated but somewhat more ideologically distinct parties (Sosialistisk Venstreparti, Fremskrittspartiet, Miljøpartiet De Grønne, Pasientfokus, Other) received a weight of 1,5 (Bjerkem, 2016; Jupskås & Langsæther, 2023)
3. The most ideologically extreme party (Rødt) was treated as fringe and given the weight of 2,0 (Center for Strategic & International Studies, 2021).

In the MNAR dataset, each respondent's MAR-based deletion weight (derived from supercluster behavior) was further multiplied by the weight associated with their chosen party. This had the effect of increasing the deletion likelihood for parties assumed to be less socially acceptable. In the MAR dataset, this MNAR-based adjustment was omitted, keeping the deletion probabilities strictly dependent on observed response patterns.

After applying these mechanisms, 20 % of the dataset was deleted:

1. 4 % of values were removed at random (MCAR),
2. 16 % were removed based on structured probabilities (either MAR or MNAR, depending on the dataset).

This resulted in two final versions of each dataset:

1. One with MCAR + MAR only missingness, for a simpler scenario.
2. One with MCAR + MAR + MNAR, for a more complex and realistic missing data scenario.

These datasets allow us to evaluate the robustness of imputation methods under both moderately and highly difficult missingness conditions.

Table 2

Original, MAR, and MNAR Distributions for Sweden: Valid Values, *%*

Party	Original	MAR	MNAR
Centern	9,6	8,9	9,5
Kristdemokraterna	5,7	6,1	6,4
Liberalerna	4,5	4,3	5,0
Miljöpartiet de gröna	6,2	6,4	6,4
Moderata samlingspartiet	19,5	18,8	18,4
Socialdemokraterna	35,3	36,1	37,1
Sverigedemokraterna	8,0	8,2	7,6
Vänsterpartiet	10,0	10,1	8,7
Annat parti	1,3	1,2	1,0

Source: *compiled by the authors.*

Table 3

Original, MAR, and MNAR Distributions for Norway: Valid Values, *%*

Party	Original	MAR	MNAR
Rødt	5,0	4,6	4,2
Sosialistisk Venstreparti	9,0	8,9	8,5
Arbeiderpartiet	27,4	27,6	27,7
Venstre	4,5	4,5	4,9
Kristelig Folkeparti	3,1	3,2	2,9
Senterpartiet	12,8	12,8	13,1
Høyre	23,9	24,5	24,8
Fremskrittspartiet	7,3	7,0	7,2
Miljøpartiet De Grønne	4,2	4,2	4,0
Other	2,8	2,8	2,6

Source: *compiled by the authors.*

As intended by our missing data simulation, subtle but systematic differences emerge in the observed party distributions when comparing the MAR and MNAR datasets to the original complete data (tables 2 and 3). For instance, in the Swedish MNAR dataset, the observed share for Socialdemokraterna increased by 1,8 percentage points compared to its original distribution. Concurrently, parties we designated as more polarizing for the MNAR simulation, such as Vänsterpartiet and Sverigedemokraterna, saw their observed shares decrease slightly, reflecting the increased likelihood of their supporters' responses being removed. Similar trends can be observed for the case of Norway, with Rødt support going down from 5,0 % in full dataset to 4,2 % in the MNAR dataset, while several mainstream parties show stable or slightly increased observed shares after the introduction of missingness.

Even these seemingly modest distributional shifts, resulting from the simulated nonresponse, necessitate the appropriate missing data handling. Failure to address such systematic missingness can compromise not only statistical power (if listwise deletion were used) but, more importantly, the validity of research findings derived from the incomplete data.

3. METHODS OF CHOICE

Handling missing categorical data poses greater methodological challenges than working with ordinal or metric variables. While many imputation methods are tailored to numeric data and can be expanded to handle ordinal data, options for nominal variables are more limited and less accessible. In this study, we adopt a multiple imputation framework, which is widely regarded as one of the most robust and theoretically grounded approaches to missing data in the social sciences (e.g.: Alwateer et al., 2024; Newman, 2014). It has been proven experimentally to be a robust approach when dealing with a significant case of missing data (Kovtun, & Fataliieva, 2020a). It also has broad software support, making it a practical and accessible choice for sociologists (Alwateer et al., 2024; Kovtun, & Fataliieva, 2020b).

We evaluate the following key imputation methods, implemented in R through the mice and VIM packages (Buuren, & Groothuis-Oudshoorn, 2011; Kowarik, & Templ, 2016):

1. Multinomial Logistic Regression

A parametric method that uses multinomial logistic regression to model relationships between variables. It is well-suited for imputing categorical outcomes with multiple, unordered levels, as is the case with "party choice". (Agresti, 2002)

2. Random Forest

A nonparametric ensemble method based on aggregating a multitude of decision trees. It captures complex, nonlinear relationships and interactions, and is known for strong performance in imputation tasks. (Stekhoven, & Bühlmann, 2012)

3. CART (Classification and Regression Trees)

A single-tree version of Random Forest. While less powerful, it is more interpretable and computationally lighter, making it attractive for quick applications. (Breiman, 1984)

4. K-Nearest Neighbors (KNN)

A simple, distance-based method that imputes missing values based on the most similar observations, and capable of handling both numerical and categorial data. (Alwateer et al., 2024)

All four model-based methods are employed within a multiple imputation (MI) framework. MI addresses a key limitation of single imputation by generating m plausible imputed values for each missing entry, creating m complete datasets (Little, & Rubin, 1989; Rubin, 1987). The differences between these m datasets represent the uncertainty surrounding the true values of the missing data. By analyzing each dataset and then pooling the results using Rubin's Rules, MI provides parameter estimates and standard errors that validly reflect this imputation uncertainty, leading to more accurate statistical inferences than methods that treat imputed values as known.

We performed multiple imputation generating m=20 imputed datasets for each scenario. The choice of m=20 is grounded in modern recommendations suggesting that the number of imputations should ideally be comparable to, or exceed, the percentage of cases with missing data on the variable being imputed to ensure stable estimates and reliable standard errors from Rubin's Rules (Graham, 2009; White et al., 2011). Given that we simulate 20 % missingness on our target variables, m=20 is a value that follows this guideline.

We also include Hot Deck imputation and Mode imputation as baseline methods. In Hot Deck approach, missing values are filled in by randomly selecting observed values from available cases (Andridge, & Little, 2010). Mode imputation imputes the most commonly observed value for all missing values. While simple to implement, these methods serve here as a contrast, showcasing the potential downsides of simpler imputation strategies, particularly when no predictive modeling is used.

More complex methods, such as Bayesian neural networks or Dirichlet Process mixture models that were suggested in some of the modern papers (Manrique-Vallier, & Reiter, 2013; Murray, & Reiter, 2016), were not touched upon. While showing theoretical promise, these approaches are either not yet accessible in standard statistical software or packages, computationally demanding, or require custom implementation and significant expertise, making them impractical to recommend for the imputation procedures in social science. In addition, as shown, for instance, in comparative study by Wongkamthong, & Akande (2020), innovative methods such as Generative Adversarial Imputation Nets did not outperform more standard approaches for preserving multivariate relationships during imputation, suggesting that these methods might not necessary carry an improvement compared to the simpler ones (Wongkamthong, & Akande, 2020).

4. PREDICTIVE MODELS

Imputing missing data requires a theoretically grounded predictive model that includes meaningful covariates. For this study, our initial selection of potential predictors for 'party choice' in Sweden and Norway was guided by established sociological and political science theories of voting behavior. We aimed for a comprehensive pool of variables from the European Social Survey (ESS) that capture key domains known to influence political preferences. A glossary defining all ESS variables referenced in this study is provided in Appendix A. These domains included:

1. Socio-demographic characteristics: Such as age and gender.
2. Economic evaluations and attitudes: Including perceptions of the national economy and views on income inequality (e.g., stfeco: "How satisfied with present state of economy in country", gincdif: "Government should reduce differences in income levels").

3. Social and cultural values: Encompassing attitudes towards immigration (e.g., imueclt: "Country's cultural life undermined or enriched by immigrants", imwbcnt: "Immigrants make country worse or better place to live"), environmental concerns (e.g., wrclmch: "How worried about climate change"), religiousness (e.g., rlgdgr: "How religious are you"), and views on equality and social norms (e.g., ipeqopta: "Important that people are treated equally and have equal opportunities", freehms: "Gays and lesbians free to live life as they wish").

4. Political trust and engagement: Covering trust in various institutions (e.g., trstprl: "Trust in country's parliament", trstplt: "Trust in politicians"), satisfaction with democracy (e.g., stfdem: "How satisfied with the way democracy works in country"), and political interest(e.g., polintr: "How interested in politics", cptppola: "Confident in own ability to participate in politics").

5. Personal values and orientations: Such as "Placement on left right scale" (lrscale) and other human values items available in the ESS that reflect broader personal motivations (e.g., ipstrgva: "Important that government is strong and ensures safety", impdiffa: "Important to try new and different things in life").

We then tested their statistical associations with the voting variable (prtvtdse/prtvtcno) using Kruskal–Wallis tests for ordinal variables, and Cramer's V for categorial variables.

When choosing a set of the predictors, one of the most common problems is to decide on whether the model should include only the key predictors, with highest significance, reducing the amount of potential noise, or should include as much variables as possible, as long as they have statistically significant correlations and are grounded in theoretical understanding. As such, for our experiment, we've decided to compare the performance of models of different size, each including the variables based on criteria of varied strictness:

1) Small model: includes only the top-10 variables based on the lowest Benjamini-Hochberg adjusted p-values.

2) Medium model: includes the top-20 variables based on the lowest Benjamini-Hochberg adjusted p-values.

3) Large model: includes all variables where Benjamini-Hochberg adjusted p-value is lower than 0,01, and either Eta-Squared (for ordinal variables) is above 0,03, or V (for categorial variables) is above 0,1.

4) Inclusive model: includes all variables where Benjamini-Hochberg adjusted p-value is lower than 0,05, and either Eta-Squared (for ordinal variables) is above 0,01, or V (for categorial variables) is above 0,05.

The specific variables included in each of these four predictor sets for both Sweden and Norway, along with their statistical properties, are detailed in the appendixes B and C respectively.

In order to ensure robust selection of the predictors, instead of using raw p-values, we employed Benjamini-Hochberg adjusted p-values, to control the False Discovery Rate, for the cases where we need to evaluate numerous potential predictors (Benjamini & Hochberg, 1995). By using adjusted p-values, we reduce the likelihood of including variables that appear statistically significant purely by chance, leading to more reliable predictor sets. Furthermore, for our "Large" and "Inclusive" models, we incorporated minimum effect size thresholds (η^2 for ordinal, V for categorical) alongside the adjusted p-value criteria. This approach ensures that included predictors demonstrate not only statistical significance but also a minimum level of practical association with the outcome variable. It prevents the inclusion of variables with statistically significant but substantively trivial relationships, which is a frequent occurrence when dealing with large sample sizes that are common in survey research like the ESS (Fritz et al., 2012).

The specific thresholds for p-values and effect sizes for the "Large" and "Inclusive" models were chosen to create a gradation in model complexity, allowing for a detailed analysis on how predictor set size impacts the process of data imputation.

For Sweden, this resulted in a 'Small' set of 10 predictors, a 'Medium' set of 20, a 'Large' set of 38 predictors, and an 'Inclusive' set of 50 predictors. For Norway, the corresponding predictor counts were 10, 20, 45, and 46. Across both countries, variables such as placement of the individual on the left-right political scale (lrscale) and attitudes towards immigration (e.g. imueclt: "Country's cultural life undermined or enriched by immigrants" and imbgeco: "Immigration bad or good for country's economy") were consistently noted as highly significant predictors and were included in most, if not all, predictor sets. As a result, we've ended up with 4 sets of predictors for Sweden and 4 sets of predictors for Norway. The

predictor sets for Sweden and Norway ended up being slightly different, although with a considerable overlap in areas like attitudes toward immigration, climate change, equality, personal values, trust in institutions, religiousness, and gender identity. These statistically significant predictors capture key dimensions that are theoretically linked to voting behavior: political trust, social attitudes, climate views, religiosity, and economic perspectives. We incorporate these variables during the procedure of multiple imputation.

After finalizing the predictor sets, we assess the inherent predictive capabilities of all of these sets of predictors by performing a 5-fold cross-validation using multinomial logistic regression to predict the original party of choice (country-specific) on the complete cases available for each set. The performance of each predictor set was evaluated using three key metrics:

- Accuracy: The proportion of correctly predicted party choices, indicating overall predictive correctness.
- Cohen's Kappa: A measure of agreement between predicted and actual party choices, corrected for the agreement expected by chance alone. This is particularly important in multi-class scenarios as it provides a more robust measure than simple accuracy.
- Macro F1-score: The unweighted average of the F1-scores (harmonic mean of precision and recall) calculated for each party. This metric is sensitive to performance across all classes, including less frequent ones.

Table 4

Preliminary Predictive Performance of Predictor Sets

Predictor Set	Country	Accuracy	Kappa	Macro F1	N
SE_Small	SE	0,52	0,35	0,43	1009
SE_Medium	SE	0,48	0,31	0,38	918
SE_Large	SE	0,44	0,29	0,36	865
SE_Inclusive	SE	0,42	0,27	0,33	844
NO_Small	NO	0,48	0,34	0,37	1005
NO_Medium	NO	0,45	0,31	0,36	980
NO_Large	NO	0,41	0,29	0,34	823
NO_Inclusive	NO	0,36	0,24	0,32	806

Note: *Metrics are means from 5-fold cross-validation. N refers to the number of complete cases used in the cross-validation for each predictor set.*

Source: *compiled by the authors.*

The "Small" predictor set, which included only the top 10 predictors, demonstrated superior performance across all evaluation metrics (Mean Accuracy, Mean Cohen's Kappa, and Mean Macro F1) in both Swedish and Norwegian samples. For instance, for the Swedish case, the Mean Kappa values exhibited a consistent decline with increasing predictor set size: 0,35 for "Small", 0.31 for "Medium", 0,29 for "Large", and 0,27 for "Inclusive".

As a preliminary hypothesis, we speculate that this performance pattern appears to be substantially influenced by the number of complete observations (N_obs) available for the analysis. The listwise deletion procedure for handling missing predictor values resulted in larger sample sizes for "Small" models (e.g. N=1009 for SE_Small) compared to their more comprehensive counterparts (e.g. N=844 for SE_Inclusive).

These initial results indicate that while expanded predictor sets may theoretically capture more information, their practical utility in direct predictive modeling may be compromised by reduced effective sample sizes due to cumulative missingness patterns. Additionally, the incorporation of multiple predictors with relatively weak associations may introduce noise and potential overfitting issues, even before considering the complexities of the imputation stage. In further analysis, we'll evaluate the performance of these predictor sets within a multiple imputation framework.

5. RESULTS

We performed the multiple imputation (m=20, as detailed in the "Methods of Choice" chapter) for the each of the following combinations:

1) 4 datasets: MAR dataset for Sweden, MNAR dataset for Sweden, MAR dataset for Norway, and MNAR dataset for Norway.

2) 4 sets of predictors for each of the countries: "Small", "Medium", "Large", and "Inclusive", which vary between Sweden and Norway.

3) 6 methods of imputation: Multinomial Logistic Regression (LogReg), Random Forest (RF), Classification and Regression Trees (CART), KNN, Hot Deck, and Mode.

To evaluate the performance of each imputation method, we compared imputed values to the original, ideal dataset, using the three metrics: Accuracy, Cohen's Kappa, and Macro F1. The detailed performance metrics for each combination of dataset, predictor set, and imputation method are presented in tables 5 through 8.

Table 5

Results for the Swedish MAR Dataset

Predictor Set	Method	Accuracy	Kappa	Macro F1
Small	KNN	0,41	0,23	0,32
Inclusive	LogReg	0,36	0,22	0,29
Large	LogReg	0,36	0,22	0,29
Small	CART	0,36	0,22	0,28
Small	RF	0,37	0,21	0,29
Small	LogReg	0,36	0,20	0,29
Medium	CART	0,34	0,19	0,27
Inclusive	RF	0,36	0,19	0,28
Medium	LogReg	0,34	0,19	0,27
Medium	KNN	0,37	0,19	0,28
Large	RF	0,35	0,18	0,28
Medium	RF	0,34	0,17	0,26
Inclusive	CART	0,33	0,17	0,27
Large	CART	0,32	0,16	0,27
Inclusive	KNN	0,35	0,15	0,25
Large	KNN	0,31	0,11	0,24
Inclusive	Mode	0,32	0,00	0,49
Large	Mode	0,32	0,00	0,49
Medium	Mode	0,32	0,00	0,49
Small	Mode	0,32	0,00	0,49
Inclusive	HotDeck	0,19	-0,01	0,17
Large	HotDeck	0,19	-0,01	0,17
Medium	HotDeck	0,19	-0,01	0,17
Small	HotDeck	0,19	-0,01	0,17

Note: *N missing values = 209, "LogReg" = Logistic Regression, "RF" = Random Forest*.
Source: *compiled by the authors.*

Table 6

Results for the Swedish MNAR Dataset

Predictor Set	Method	Accuracy	Kappa	Macro F1
1	2	3	4	5
Inclusive	LogReg	0,37	0,23	0,31
Small	RF	0,37	0,22	0,32
Large	LogReg	0,35	0,22	0,30
Large	CART	0,35	0,21	0,28
Small	KNN	0,37	0,21	0,31
Inclusive	CART	0,35	0,21	0,29
Small	LogReg	0,35	0,21	0,31
Medium	LogReg	0,35	0,20	0,30
Small	CART	0,35	0,20	0,30
Medium	RF	0,36	0,20	0,30

The End of theTable 6

1	2	3	4	5
Medium	CART	0,33	0,19	0,28
Large	RF	0,35	0,18	0,27
Medium	KNN	0,35	0,18	0,26
Inclusive	RF	0,34	0,17	0,29
Inclusive	KNN	0,33	0,14	0,25
Large	KNN	0,30	0,12	0,23
Inclusive	HotDeck	0,19	0,00	0,17
Large	HotDeck	0,19	0,00	0,17
Medium	HotDeck	0,19	0,00	0,17
Small	HotDeck	0,19	0,00	0,17
Inclusive	Mode	0,28	0,00	0,44
Large	Mode	0,28	0,00	0,44
Medium	Mode	0,28	0,00	0,44
Small	Mode	0,28	0,00	0,44

Note: *N missing values = 213, "LogReg" = Logistic Regression, "RF" = Random Forest.*
Source: *compiled by the authors.*

Table 7

Results for the Norwegian MAR dataset

Predictor Set	Method	Accuracy	Kappa	Macro F1
Inclusive	LogReg	0,34	0,22	0,31
Large	LogReg	0,34	0,22	0,30
Medium	LogReg	0,33	0,19	0,30
Small	KNN	0,34	0,18	0,29
Medium	KNN	0,34	0,18	0,32
Large	CART	0,30	0,17	0,25
Inclusive	CART	0,30	0,16	0,25
Small	RF	0,31	0,16	0,26
Small	LogReg	0,30	0,16	0,27
Medium	RF	0,30	0,15	0,27
Medium	CART	0,29	0,15	0,26
Large	KNN	0,31	0,14	0,31
Inclusive	RF	0,30	0,14	0,26
Large	RF	0,29	0,14	0,27
Small	CART	0,28	0,14	0,24
Inclusive	KNN	0,28	0,11	0,27
Inclusive	HotDeck	0,17	0,00	0,17
Large	HotDeck	0,17	0,00	0,17
Medium	HotDeck	0,17	0,00	0,17
Small	HotDeck	0,17	0,00	0,17
Inclusive	Mode	0,27	0,00	0,42
Large	Mode	0,27	0,00	0,42
Medium	Mode	0,27	0,00	0,42
Small	Mode	0,27	0,00	0,42

Note: *N missing values = 207, "LogReg" = Logistic Regression, "RF" = Random Forest.*
Source: *compiled by the authors.*

Table 8

Results for the Norwegian MNAR Dataset

Predictor Set	Method	Accuracy	Kappa	Macro F1
Large	LogReg	0,32	0,20	0,29
Inclusive	LogReg	0,31	0,19	0,27
Small	CART	0,32	0,19	0,28
Large	CART	0,31	0,19	0,28
Small	KNN	0,34	0,19	0,27
Medium	LogReg	0,31	0,18	0,29
Inclusive	CART	0,31	0,18	0,27
Medium	KNN	0,33	0,17	0,29
Small	LogReg	0,30	0,17	0,27
Small	RF	0,30	0,17	0,27
Medium	RF	0,31	0,16	0,26
Large	RF	0,30	0,15	0,26
Inclusive	RF	0,30	0,15	0,25
Medium	CART	0,29	0,14	0,28
Large	KNN	0,27	0,10	0,28
Inclusive	KNN	0,26	0,09	0,28
Inclusive	Mode	0,26	0,00	0,42
Large	Mode	0,26	0,00	0,42
Medium	Mode	0,26	0,00	0,42
Small	Mode	0,26	0,00	0,42
Inclusive	HotDeck	0,16	-0,01	0,16
Large	HotDeck	0,16	-0,01	0,16
Medium	HotDeck	0,16	-0,01	0,16
Small	HotDeck	0,16	-0,01	0,16

Note: *N missing values = 206, "LogReg" = Logistic Regression, "RF" = Random Forest.*
Source: *compiled by the authors.*

Our findings include:

1. First of all, imputation of missing categorial data when dealing with party preferences in a field with many (10+) parties is a challenging task, regardless of the method chosen. Political preferences are complex and even with many predictors, capturing the nuances perfectly is challenging.

2. However, the more sophisticated model-based methods have shown results that are superior to the more "naïve" methods of handling missing data such as Hot Deck and Mode imputation, and as such we strongly advice employment of said methods when dealing with a substantial case of missing data over the more simplistic ones.

3. As expected, Hot Deck imputation consistently produced the lowest accuracy and near-zero or even negative Kappa values. Mode imputation, while achieving comparatively high F1 Macro score, yielded Kappa values of 0, which indicated no improvements over purely random attribution of values. The F1 Macro score can be explained by the calculation of the metric in imbalanced scenarios: high value stems from a single class (most frequent party), masking poor performance on other classes. This reinforces the need to avoid simplistic, non-model-based approaches when handling categorical data.

4. Across all methods and missingness types, the Swedish datasets yielded slightly higher overall performance. This suggests that the underlying mechanism of missingness in Sweden may be less complex or more predictable, allowing imputation models to better capture the patterns.

5. Generally, results revealed marginally superior performance of imputation models under MAR conditions versus MNAR ones, as quantified by Mean Kappa metrics, which was expected based on theoretical foundation. However, the difference was not particularly sizeable across the top-performing methods. This could suggest that either the simulated MNAR mechanism, while present, was not sufficiently strong to drastically alter relative method performance, or that the leading methods (particularly Multinomial Logistic Regression) possess a degree of robustness, potentially by leveraging informative predictors that also correlate with the missingness mechanism.

6. While during the preliminary evaluation smaller sets of predictors exhibited higher scores, which we proposed could be tied to the larger N, under the circumstances of the real case of missing data, “Large” and “Inclusive” models generally showcased slightly higher results than their smaller counterparts. The notable exception is KNN, which performed better under smaller sets of predictors, which reflects the complexity of choosing closest neighbor in higher dimensional spaces introduced by extra predictors. For methods such as Multinomial Logistic Regression, this reinforces the idea that the models should try to include a wide set of statistically significant and theoretically sound predictors when using a robust multiple imputation framework like mice which can handle missingness within the predictor set.

7. Overall, Multinomial Logistic Regression, provided by mice's polyreg, consistently delivers the best or near-best performance across different scenarios, predictor set sizes, and missingness mechanisms.

8. KNN has emerged as a method that occasionally challenged the Multinomial Logistic Regression’s results, and while it did not consistently outperform it, the fact that it performs better at the lower amount of predictors is a notable finding that could make it applicable in situations when usage of a large model is deemed unfeasible computationally, or impractical otherwise.

9. While tree-based methods (Random Forest, CART) provided for comparable performance, overall, they did not consistently outperform the Multinomial Logistic Regression methods, and as such may be less fitting for the goal.

Beyond imputation quality, practical application also considers computational efficiency. We therefore compared the runtimes for each method and predictor set combination (m=20), with an illustrative example for the SE_MAR scenario presented in Table 9 (full timings in Appendix C).

Table 9

Illustrative Computational Time (Seconds) for Imputation Methods (m=20) in the SE_MAR Scenario

Dataset	Predictor Set	Method	Duration (s)
SE_MAR	Small	LogReg	39
SE_MAR	Small	RF	24
SE_MAR	Small	CART	15
SE_MAR	Small	KNN	9
SE_MAR	Medium	LogReg	128
SE_MAR	Medium	RF	85
SE_MAR	Medium	CART	80
SE_MAR	Medium	KNN	23
SE_MAR	Large	LogReg	1094
SE_MAR	Large	RF	937
SE_MAR	Large	CART	941
SE_MAR	Large	KNN	55
SE_MAR	Inclusive	LogReg	1519
SE_MAR	Inclusive	RF	1283
SE_MAR	Inclusive	CART	1296
SE_MAR	Inclusive	KNN	86

Note: *“LogReg” = Logistic Regression, “RF” = Random Forest. Hot Deck and Mode imputation times were <1s and were omitted for brevity. Full table of computational times is available in the appendix C.*

Source: *compiled by the authors.*

While timings for the “Small” sets were negligible, as expected, with the increase of the predictor set size increased, computation times rose substantially. For Multinomial Logistic Regression, moving from the “Small” (39s) to the “Inclusive” set (1519s) resulted in an approximately 39-fold increase in runtime. Similar steep increases were observed for Random Forest and CART (e.g., with “Inclusive” CART imputation taking 1283s). These timings are for m=20 imputations and would scale proportionally with the

different amount of imputations. Sociologists facing computational constraints should consider this trade-off. While imputation using LogReg with larger predictor sets often yielded strong performance in our study, the associated computational cost was substantially greater. In contrast, methods like KNN with smaller predictor sets offered a much faster alternative, though with more variable imputation quality depending on the scenario.

These findings on both imputation quality and computational cost provide a comprehensive basis for the practical recommendations and theoretical implications discussed in the subsequent chapter.

6. CONCLUSIONS

This study addressed the common challenge of handling missing categorical data in quantitative sociology, a task complicated by the unsuitability of simple approaches such as listwise deletion, mean imputation, or hot-deck, and the nuanced nature of variables like political party preference. Through a comprehensive simulation study involving sophisticated missing data mechanisms (MCAR, MAR, and MNAR) and systematically varied predictor set complexities on European Social Survey data from Sweden and Norway, we aimed to construct empirically grounded guidelines on the selection and application of multiple imputation methods for such data. While derived from the specific context of political preference data, the principles resulting from our findings offer valuable guidelines for researchers across a range of sociological subfields encountering similar challenges with missing categorical variables.

Our investigation offers several key insights for sociological researchers dealing with missing categorical responses. First of all, we demonstrate that while the imputation of multi-category nominal variables remains a complex endeavor, sophisticated model-based approaches offer significant improvements over simplistic methods, with multinomial logistic regression emerging as a particularly robust and accessible strategy. Furthermore, our findings highlight the intricate relationship between predictor set size, computational resources, and imputation efficacy, offering practical considerations for model specification.

The key findings from our evaluation are as follows:

1) The task of imputing party choice in these multi-party systems proved challenging, with even the best-performing methods achieving modest performance metrics. This showcases the inherent complexity of the political attitudes and difficulties regarding their prediction.

2) As hypothesized, model-based imputation techniques significantly outperformed naïve methods. Both Mode and Hot Deck imputation resulted in near-zero Cohen’s Kappa values, highlighting their inadequacy for preserving meaningful information and relationships in the data. This strongly advocates for the adoption of more principled imputation approaches in sociological research, especially when the fraction of missing data is non-negligible.

3) With regards to the choice of imputation algorithm, multinomial logistic regression (implemented via mice’s polyreg function) consistently emerged as the most robust and often highest-performing method across the various scenarios, countries, and missingness mechanisms.

4) The impact of predictor set size revealed a nuanced picture. While our preliminary direct predictive modeling suggested an advantage for smaller predictor sets (due to maximizing complete cases), the main imputation results indicated that multinomial logistic regression, in particular, often benefited from larger predictor sets. This suggests that robust MI frameworks like mice can effectively leverage the information from a wider array of predictors by managing their internal missingness. However, this benefit was not universal across all methods; K-Nearest Neighbors (KNN) imputation, for instance, performed better with smaller predictor sets, consistent with its known sensitivity to high dimensionality.

Based on these findings, we suggest several practical recommendations for sociologists. Multinomial logistic regression emerges as a strong default choice for imputing categorical variables like party preference, especially when a comprehensive set of predictors is available. Its consistent performance and availability in accessible software like R’s mice package make it a practical option that can be recommended as a robust starting option for complex cases of missing data. For situations demanding high computational efficiency or when only a very small set of strong predictors can be reliably used, K-Nearest Neighbors (KNN) imputation may serve as a viable, faster alternative, though its performance can degrade with larger, more complex predictor sets. While tree-based methods like Random Forest and CART are deemed to be

powerful and relatively accessible ML-based methods of handling missing data, they did not consistently outperform multinomial logistic regression in our simulations.

An often-overlooked practical consideration is computational cost. Our analysis of runtimes (see Appendix C for full details) revealed that while using larger predictor sets with methods like multinomial logistic regression often improved the imputation quality, the computational burden increased dramatically—for example, from approximately 39 seconds for a "Small" set to over 1500 seconds for an "Inclusive" set with LogReg (m=20). Researchers must weigh the potential for marginal gains in imputation accuracy against these substantial increases in computation time, especially in projects with limited resources or very large datasets.

Our study, while providing valuable insights, has several limitations that could suggest the direction for future research. First, our findings were based off the simulation of missingness for the variable "party voted" for two countries, Sweden and Norway. While this represents a common scenario of a significant missing data in the quantitative sociology, whether the findings can be generalized to the other spheres of sociological research requires further investigation. Secondly, the real-world MNAR could be more complex compared to the one we've proposed in the simulation. Future studies could explore a wider range of MNAR scenarios. Finally, although our study focused on practical, accessible methods, a range of more advanced techniques exist, such as Bayesian models or neural networks. Currently, their higher computational demands, implementation complexity, and often reduced interpretability limit their adoption in quantitative sociological research. As these approaches become more standardized, computationally efficient, and user-friendly, they may eventually offer powerful new options for handling missing categorical data and warrant their inclusions in comparative studies that pursue the goal of providing practical recommendations to the researchers.

REFERENCES

Agresti, A. (2002). *Categorical Data Analysis* (1st ed.). Wiley. https://doi.org/10.1002/0471249688

Alwateer, M., Atlam, E.-S., El-Raouf, M. M. A., Ghoneim, O. A., & Gad, I. (2024). Missing Data Imputation: A Comprehensive Review. *Journal of Computer and Communications*, *12*(11), 53–75. https://doi.org/10.4236/jcc.2024.1211004

Andridge, R. R., & Little, R. J. A. (2010). A Review of Hot Deck Imputation for Survey Non-response. *International Statistical Review*, *78*(1), 40–64. https://doi.org/10.1111/j.1751-5823.2010.00103.x

Bakker, R., Hooghe, L., Jolly, S., Marks, G., Polk, J., Rovny, J., Steenbergen, M., & Anna Vachudova, M. (2020). *2019 Chapel Hill Expert Expert Survey (CHES)* [Dataset]. https://www.chesdata.eu/2019-chapel-hill-expert-survey

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

Bjerkem, J. (2016). The Norwegian Progress Party: An established populist party. *European View*, *15*(2), 233–243. https://doi.org/10.1007/s12290-016-0404-8

Breiman, L. (1984). *Classification and Regression Trees*. Wadsworth International Group.

Bulent, K. (2020). *The Sweden Democrats: Killer of Swedish Exceptionalism*. European Center for Populism Studies (ECPS). https://doi.org/10.55271/op0001

Buuren, S. V., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in *R*. *Journal of Statistical Software*, *45*(3). https://doi.org/10.18637/jss.v045.i03

Center for Strategic & International Studies (2021). *European Election Watch: Norway 2021*. Center for Strategic & International Studies. Retrieved May 02, 2025 from https://www.csis.org/programs/europe-russia-and-eurasia-program/projects/european-election-watch/2021-elections/norway

Dong, W., Fong, D. Y. T., Yoon, J., Wan, E. Y. F., Bedford, L. E., Tang, E. H. M., & Lam, C. L. K. (2021). Generative adversarial networks for imputing missing data for big data clinical research. *BMC Medical Research Methodology*, *21*(1), 78. https://doi.org/10.1186/s12874-021-01272-3

ESS ERIC (2024). *ESS11—Integrated file, edition 2.0* [Dataset]. Sikt – Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/ESS11E02_0

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology. General*, *141*(1), 2–18. https://doi.org/10.1037/a0024338

Ge, Y., Li, Z., & Zhang, J. (2023). A simulation study on missing data imputation for dichotomous variables using statistical and machine learning methods. *Scientific Reports*, *13*(1), 9432. https://doi.org/10.1038/s41598-023-36509-2

Graham, J. W. (2009). Missing Data Analysis: Making It Work in the Real World. *Annual Review of Psychology*, 60(1), 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530

Jupskås, A. R., & Langsæther, P. E. (2023). Norway. In F. Escalona, D. Keith, & L. March (Eds.), *The Palgrave Handbook of Radical Left Parties in Europe* (pp. 423–447). Palgrave Macmillan UK. https://doi.org/10.1057/978-1-137-56264-7_15

Kovtun, N. V., & Fataliieva, A.-N. Ya. (2020a). New Trends in Evidence-based Statistics: Data Imputation Problems. *Statistics of Ukraine*, *87*(4), 4–13. https://doi.org/10.31767/su.4(87)2019.04.01

Kovtun, N. V., & Fataliieva, A.-N. Ya. (2020b). Software Implementation of Missing Data Recovery: Comparative Analysis. *Statistics of Ukraine*, *91*(4), 12–20. https://doi.org/10.31767/su.4(91)2020.04.02

Kowarik, A., & Templ, M. (2016). Imputation with the *R* Package VIM. *Journal of Statistical Software*, 74(7). https://doi.org/10.18637/jss.v074.i07

Lang, K. M., & Wu, W. (2017). A Comparison of Methods for Creating Multiple Imputations of Nominal Variables. *Multivariate Behavioral Research*, 52(3), 290–304. https://doi.org/10.1080/00273171.2017.1289360

Lee, J. H., & Huber, J. C. (2021). Evaluation of Multiple Imputation with Large Proportions of Missing Data: How Much Is Too Much? *Iranian Journal of Public Health*. https://doi.org/10.18502/ijph.v50i7.6626

Little, R. J. A., & Rubin, D. B. (1989). The Analysis of Social Science Data with Missing Values. *Sociological Methods & Research*, 18(2–3), 292–326. https://doi.org/10.1177/0049124189018002004

Manrique-Vallier, D., & Reiter, J. P. (2013). *Bayesian multiple imputation for large-scale categorical data with structural zeros*. https://hdl.handle.net/1813/34889

Murray, J. S., & Reiter, J. P. (2016). Multiple Imputation of Missing Categorical and Continuous Values via Bayesian Mixture Models with Local Dependence. *Journal of the American Statistical Association*, *111*(516), 1466–1479. https://doi.org/10.1080/01621459.2016.1174132

Newman, D. A. (2014). Missing Data: Five Practical Guidelines. *Organizational Research Methods*, *17*(4), 372–411. https://doi.org/10.1177/1094428114548590

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. https://doi.org/10.1093/biomet/63.3.581

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys* (1st ed.). Wiley. https://doi.org/10.1002/9780470316696

Stekhoven, D. J., & Bühlmann, P. (2012). MissForest—Non-parametric missing value imputation for mixed-type data. *Bioinformatics*, *28*(1), 112–118. https://doi.org/10.1093/bioinformatics/btr597

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, *30*(4), 377–399. https://doi.org/10.1002/sim.4067

Wongkamthong, C., & Akande, O. (2020). *A Comparative Study of Imputation Methods for Multivariate Ordinal Data*. https://doi.org/10.48550/ARXIV.2010.10471

APPENDIX

Appendix A

Full Variable Glossary

Variable	Description	Used for Imputation	Used for Clusterization
1	**2**	**3**	**4**
actrolga	Able to take active role in political group	✓	✓
agea	Age of respondent, calculated	✓	
impcntr	Allow many/few immigrants from poorer countries outside Europe	✓	✓
imdfetn	Allow many/few immigrants of different race/ethnic group from majority	✓	✓
imsmetn	Allow many/few immigrants of same race/ethnic group as majority	✓	✓
hmsfmlsh	Ashamed if close family member gay or lesbian	✓	✓
eqmgmbg	Bad or good for businesses in [country] if equal numbers of women and men are in higher management positions		✓
eqpaybg	Bad or good for economy in [country] if women and men receive equal pay for doing the same work		✓
eqwrkbg	Bad or good for family life in [country] if equal numbers of women and men are in paid work		✓

The Continuation of the Appendix A

1	2	3	4
eqpolbg	Bad or good for politics in [country] if equal numbers of women and men are in positions of political leadership		✓
rlgblg	Belonging to particular religion or denomination		✓
ccnthum	Climate change caused by natural processes, human activity, or both	✓	✓
cptppola	Confident in own ability to participate in politics	✓	✓
loylead	Country needs most loyalty towards its leaders		✓
imueclt	Country's cultural life undermined or enriched by immigrants	✓	✓
eqparep	Dividing the number of seats in parliament equally between women and men		✓
euftf	European Union: European unification go further or gone too far		✓
freinsw	Firing employees who make insulting comments directed at women in the workplace		✓
anctrya1	First ancestry, European Standard Classification of Cultural and Ethnic Groups	✓	
hmsacld	Gay and lesbian couples right to adopt children	✓	✓
freehms	Gays and lesbians free to live life as they wish	✓	✓
gndr	Gender	✓	
gincdif	Government should reduce differences in income levels	✓	✓
edlveno	Highest level of education, Norway	✓	
edlvdse	Highest level of education, Sweden	✓	
prtdgcl	How close to party		✓
atchctr	How emotionally attached to [country]		✓
atcherp	How emotionally attached to Europe		✓
trplcnt	How fair the police in [country] treat women/men		✓
femifel	How feminine respondent feels	✓	✓
impbemw	How important being a man/woman is to the way respondent think about him/herself	✓	✓
polintr	How interested in politics	✓	✓
mascfel	How masculine respondent feels	✓	✓
rlgatnd	How often attend religious services apart from special occasions	✓	✓
pray	How often pray apart from at religious services	✓	✓
wexashr	How often women exaggerate claims of sexual harassment in the workplace		✓
weasoff	How often women get easily offended?		✓
wlespdm	How often women paid less than men for same work in [country]		✓
wsekpwr	How often women seek to gain power by getting control over men		✓
rlgdgr	How religious are you	✓	✓
stflife	How satisfied with life as a whole	✓	✓
stfeco	How satisfied with present state of economy in country	✓	✓
stfgov	How satisfied with the national government	✓	✓
stfdem	How satisfied with the way democracy works in country	✓	✓
wrclmch	How worried about climate change	✓	✓
actcomp	I act compassionately towards others, to what extent	✓	✓

The Continuation of the Appendix A

1	2	3	4
sothnds	I am sensitive to others' needs	✓	✓
liklead	I like to be a leader, to what extent	✓	✓
likrisk	I like to take risks, to what extent	✓	✓
imwbcnt	Immigrants make country worse or better place to live	✓	✓
imbgeco	Immigration bad or good for country's economy	✓	✓
ipstrgva	Important that government is strong and ensures safety	✓	
ipeqopta	Important that people are treated equally and have equal opportunities	✓	
ipmodsta	Important to be humble and modest, not draw attention	✓	
iplylfra	Important to be loyal to friends and devote to people close	✓	
impricha	Important to be rich, have money and expensive things	✓	
ipsucesa	Important to be successful and that people recognise achievements	✓	
ipbhprpa	Important to behave properly	✓	
impenva	Important to care for nature and environment	✓	
ipfrulea	Important to do what is told and follow rules	✓	
imptrada	Important to follow traditions and customs	✓	
iprspota	Important to get respect from others	✓	
ipgdtima	Important to have a good time	✓	
iphlppla	Important to help people and care for others well-being	✓	
impsafea	Important to live in secure and safe surroundings	✓	
impfreea	Important to make own decisions and be free	✓	
ipadvnta	Important to seek adventures and have an exciting life	✓	
impfuna	Important to seek fun and things that give pleasure	✓	
ipshabta	Important to show abilities and be admired	✓	
ipcrtiva	Important to think new ideas and being creative	✓	
impdiffa	Important to try new and different things in life	✓	
ipudrsta	Important to understand different people	✓	
netustm	Internet use, how much time on typical day, in minutes		✓
netusoft	Internet use, how often		✓
fineqpy	Making businesses pay a fine when they pay men more than women for doing the same work		✓
pplhlp	Most of the time people helpful or mostly looking out for themselves		✓
ppltrst	Most people can be trusted or you can't be too careful		✓
pplfair	Most people try to take advantage of you, or try to be fair		✓
nwspol	News about politics and current affairs, watching, reading or listening, in minutes		✓
nobingnd	Non-binary gender: which option respondent says best describes them	✓	✓
lrnobed	Obedience and respect for authority most important virtues children should learn		✓
lrscale	Placement on left right scale	✓	✓
psppsgva	Political system allows people to have a say in what government does	✓	✓
psppipla	Political system allows people to have influence on politics	✓	✓
eqparlv	Require both parents to take equal periods of paid leave to care for their child		✓

The End of the Appendix A

1	2	3	4
stfedu	State of education in country nowadays	✓	✓
stfhlth	State of health services in country nowadays	✓	✓
ccrdprs	To what extent feel personal responsibility to reduce climate change	✓	✓
trstprl	Trust in country's parliament	✓	✓
trstprt	Trust in political parties	✓	✓
trstplt	Trust in politicians	✓	✓
trstep	Trust in the European Parliament	✓	✓
trstlgl	Trust in the legal system	✓	✓
trstplc	Trust in the police	✓	✓
trstun	Trust in the United Nations	✓	✓
vote	Voted last national election		✓
wprtbym	Women should be protected by men		✓
wbrgwrm	Women tend to have better sense of what is right and wrong compared with men		✓
trwkcnt	Women/men: treated equally fairly in hiring, pay or promotions at work in [country]		✓
trmdcnt	Women/men: treated equally fairly when seeking medical treatment in [country]		✓
vteubcmb	Would vote for [country] to become member of European Union or remain outside	✓	
vteurmmb	Would vote for [country] to remain member of European Union or leave	✓	✓

Source: *compiled by the authors.*

Appendix B

Predictor Selection Results and Final Model Composition: Sweden

Variable	Type	N_obs	Effect Size (η^2 or V)	Adj. P	Small	Medium	Large	Inclusive
1	2	3	4	5	6	7	8	9
lrscale	H	1068	0,58	< .001	✓	✓	✓	✓
imueclt	H	1067	0,21	< .001	✓	✓	✓	✓
imbgeco	H	1061	0,20	< .001	✓	✓	✓	✓
imwbcnt	H	1071	0,19	< .001	✓	✓	✓	✓
stfgov	H	1055	0,19	< .001	✓	✓	✓	✓
gincdif	H	1070	0,18	< .001	✓	✓	✓	✓
wrclmch	H	1077	0,16	< .001	✓	✓	✓	✓
impcntr	H	1059	0,16	< .001	✓	✓	✓	✓
imdfetn	H	1059	0,12	< .001	✓	✓	✓	✓
imsmetn	H	1061	0,09	< .001	✓	✓	✓	✓
ccrdprs	H	1073	0,08	< .001		✓	✓	✓
trstep	H	1017	0,09	< .001		✓	✓	✓
ccnthum	V	1073	0,18	< .001		✓	✓	✓
hmsacld	H	1071	0,08	< .001		✓	✓	✓

The Continuation of the Appendix B

1	2	3	4	5	6	7	8	9
trstlgl	H	1073	0,07	< .001		✓	✓	✓
psppipla	H	1065	0,07	< .001		✓	✓	✓
ipeqopta	H	1051	0,07	< .001		✓	✓	✓
impenva	H	1057	0,06	< .001		✓	✓	✓
trstun	H	1043	0,06	< .001		✓	✓	✓
trstplc	H	1074	0,05	< .001		✓	✓	✓
vteurmmb	V	1061	0,15	< .001			✓	✓
hmsfmlsh	H	1071	0,05	< .001			✓	✓
freehms	H	1076	0,05	< .001			✓	✓
agea	H	1078	0,05	< .001			✓	✓
trstplt	H	1070	0,05	< .001			✓	✓
trstprl	H	1076	0,04	< .001			✓	✓
psppsgva	H	1071	0,04	< .001			✓	✓
stfdem	H	1065	0,04	< .001			✓	✓
mascfel	H	1061	0,04	< .001			✓	✓
femifel	H	1059	0,04	< .001			✓	✓
gndr	V	1078	0,19	< .001			✓	✓
nobingnd	V	1073	0,16	< .001			✓	✓
trstprt	H	1070	0,04	< .001			✓	✓
edlvdse	V	1066	0,17	< .001			✓	✓
ipfrulea	H	1051	0,03	< .001			✓	✓
actrolga	H	1074	0,03	< .001			✓	✓
impricha	H	1056	0,03	< .001			✓	✓
impbemw	H	1046	0,03	< .001			✓	✓
cptppola	H	1069	0,03	< .001				✓
imptrada	H	1054	0,03	< .001				✓
rlgdgr	H	1075	0,03	< .001				✓
stfeco	H	1068	0,02	0,002				✓
ipstrgva	H	1044	0,02	0,004				✓
pray	H	1075	0,02	0,006				✓
liklead	H	1077	0,02	0,008				✓
iprspota	H	1054	0,02	0,009				✓
rlgatnd	H	1077	0,02	0,010				✓
stfedu	H	1051	0,02	0,010				✓
ipshabta	H	1055	0,02	0,031				✓
actcomp	H	1073	0,02	0,048				✓
ipbhprpa	H	1055	0,02	0,060				
ipudrsta	H	1056	0,01	0,075				
likrisk	H	1075	0,01	0,075				
ipsucesa	H	1052	0,01	0,226				

The End of the Appendix B

1	2	3	4	5	6	7	8	9
polintr	H	1078	0,01	0,297				
stfhlth	H	1070	0,01	0,314				
ipadvnta	H	1056	0,01	0,364				
anctrya1	V	1071	0,09	0,503				
impdiffa	H	1053	0,01	0,517				
impfreea	H	1057	0,01	0,576				
ipcrtiva	H	1056	0,01	0,707				
impfuna	H	1054	0,01	0,759				
stflife	H	1075	0,01	0,768				
iphlppla	H	1057	0,01	0,788				
iplylfra	H	1056	0,00	0,788				
ipmodsta	H	1055	0,00	0,814				
impsafea	H	1057	0,00	0,815				
sothnds	H	1074	0,00	0,830				
ipgdtima	H	1058	0,00	0,849				

Notes: *Type: H = Kruskal-Wallis H, V = Cramer's V. Adj. P: Benjamini-Hochberg adjusted p-values. Values below .001 are displayed as "< .001".*

Source: *compiled by the authors.*

Appendix C

Predictor Selection Results and Final Model Composition: Norway

Variable	Type	N_obs	Effect Size (η^2 or V)	Adj. P	Small	Medium	Large	Inclusive
1	2	3	4	5	6	7	8	9
lrscale	H	1031	0,55	< .001	✓	✓	✓	✓
gincdif	H	1036	0,21	< .001	✓	✓	✓	✓
wrclmch	H	1037	0,14	< .001	✓	✓	✓	✓
imwbcnt	H	1025	0,14	< .001	✓	✓	✓	✓
stfgov	H	1034	0,13	< .001	✓	✓	✓	✓
imueclt	H	1035	0,13	< .001	✓	✓	✓	✓
ccnthum	V	1037	0,22	< .001	✓	✓	✓	✓
rlgdgr	H	1036	0,11	< .001	✓	✓	✓	✓
imbgeco	H	1027	0,11	< .001	✓	✓	✓	✓
impcntr	H	1034	0,10	< .001	✓	✓	✓	✓
rlgatnd	H	1034	0,10	< .001		✓	✓	✓
pray	H	1035	0,10	< .001		✓	✓	✓
imdfetn	H	1031	0,10	< .001		✓	✓	✓
trstprl	H	1034	0,10	< .001		✓	✓	✓
hmsacld	H	1034	0,09	< .001		✓	✓	✓
stfdem	H	1034	0,08	< .001		✓	✓	✓

The Continuation of the Appendix C

1	2	3	4	5	6	7	8	9
ccrdprs	H	1037	0,08	< .001		✓	✓	✓
trstprt	H	1034	0,08	< .001		✓	✓	✓
impenva	H	1029	0,08	< .001		✓	✓	✓
trstplt	H	1034	0,08	< .001		✓	✓	✓
edlveno	V	1036	0,16	< .001			✓	✓
psppsgva	H	1032	0,08	< .001			✓	✓
freehms	H	1034	0,08	< .001			✓	✓
psppipla	H	1032	0,08	< .001			✓	✓
vteubcmb	V	990	0,17	< .001			✓	✓
trstep	H	921	0,08	< .001			✓	✓
imsmetn	H	1030	0,06	< .001			✓	✓
actrolga	H	1034	0,05	< .001			✓	✓
hmsfmlsh	H	1030	0,05	< .001			✓	✓
agea	H	1036	0,05	< .001			✓	✓
femifel	H	1035	0,05	< .001			✓	✓
trstlgl	H	1034	0,05	< .001			✓	✓
trstun	H	1023	0,05	< .001			✓	✓
impricha	H	1030	0,05	< .001			✓	✓
gndr	V	1037	0,21	< .001			✓	✓
ipeqopta	H	1028	0,05	< .001			✓	✓
trstplc	H	1033	0,04	< .001			✓	✓
ipbhprpa	H	1028	0,04	< .001			✓	✓
imptrada	H	1032	0,04	< .001			✓	✓
stfeco	H	1032	0,04	< .001			✓	✓
nobingnd	V	1034	0,17	< .001			✓	✓
mascfel	H	1035	0,04	< .001			✓	✓
ipmodsta	H	1029	0,03	< .001			✓	✓
ipadvnta	H	1028	0,03	< .001			✓	✓
ipfrulea	H	1026	0,03	< .001			✓	✓
likrisk	H	1036	0,03	< .001				✓
ipstrgva	H	1027	0,03	0,001				✓
impbemw	H	1019	0,03	0,002				✓
ipudrsta	H	1029	0,02	0,007				✓
ipgdtima	H	1027	0,02	0,009				✓
stfedu	H	1026	0,02	0,011				✓
ipcrtiva	H	1028	0,02	0,022				✓
stflife	H	1034	0,02	0,024				✓
sothnds	H	1034	0,02	0,040				✓
ipshabta	H	1030	0,02	0,046				✓
liklead	H	1033	0,02	0,047				✓

The End of the Appendix C

1	2	3	4	5	6	7	8	9
impsafea	H	1031	0,02	0,053				
ipsucesa	H	1027	0,02	0,053				
cptppola	H	1036	0,02	0,063				
polintr	H	1037	0,01	0,106				
impdiffa	H	1029	0,01	0,107				
stfhlth	H	1036	0,01	0,148				
iphlppla	H	1031	0,01	0,180				
impfuna	H	1025	0,01	0,188				
iplylfra	H	1030	0,01	0,245				
impfreea	H	1028	0,01	0,309				
actcomp	H	1037	0,01	0,322				
iprspota	H	1025	0,01	0,364				
anctrya1	V	1034	0,16	0,746				

Notes: *Type: H = Kruskal-Wallis H, V = Cramer's V. Adj. P: Benjamini-Hochberg adjusted p-values. Values below .001 are displayed as "< .001".*

Source: *compiled by the authors.*

Appendix D

Full Timing for the Imputation Computations

Dataset	Predictor Set	Method	Duration (s)		Dataset	Predictor Set	Method	Duration (s)
1	2	3	4	5	6	7	8	9
SE_MAR	Small	LogReg	39		NO_MAR	Small	LogReg	49
SE_MAR	Small	RF	24		NO_MAR	Small	RF	24
SE_MAR	Small	CART	15		NO_MAR	Small	CART	16
SE_MAR	Small	KNN	9		NO_MAR	Small	KNN	7
SE_MAR	Small	HotDeck	0		NO_MAR	Small	HotDeck	0
SE_MAR	Small	Mode	0		NO_MAR	Small	Mode	0
SE_MAR	Medium	LogReg	128		NO_MAR	Medium	LogReg	92
SE_MAR	Medium	RF	85		NO_MAR	Medium	RF	43
SE_MAR	Medium	CART	80		NO_MAR	Medium	CART	38
SE_MAR	Medium	KNN	23		NO_MAR	Medium	KNN	15
SE_MAR	Medium	HotDeck	0		NO_MAR	Medium	HotDeck	0
SE_MAR	Medium	Mode	0		NO_MAR	Medium	Mode	0
SE_MAR	Large	LogReg	1094		NO_MAR	Large	LogReg	659
SE_MAR	Large	RF	937		NO_MAR	Large	RF	462
SE_MAR	Large	CART	941		NO_MAR	Large	CART	469
SE_MAR	Large	KNN	55		NO_MAR	Large	KNN	54
SE_MAR	Large	HotDeck	0		NO_MAR	Large	HotDeck	0
SE_MAR	Large	Mode	0		NO_MAR	Large	Mode	0
SE_MAR	Inclusive	LogReg	1519		NO_MAR	Inclusive	LogReg	855

The End of the Appendix D

1	2	3	4	5	6	7	8	9
SE_MAR	Inclusive	RF	1283		NO_MAR	Inclusive	RF	614
SE_MAR	Inclusive	CART	1296		NO_MAR	Inclusive	CART	632
SE_MAR	Inclusive	KNN	86		NO_MAR	Inclusive	KNN	82
SE_MAR	Inclusive	HotDeck	0		NO_MAR	Inclusive	HotDeck	0
SE_MAR	Inclusive	Mode	0		NO_MAR	Inclusive	Mode	0
SE_MNAR	Small	LogReg	41		NO_MNAR	Small	LogReg	49
SE_MNAR	Small	RF	24		NO_MNAR	Small	RF	23
SE_MNAR	Small	CART	15		NO_MNAR	Small	CART	15
SE_MNAR	Small	KNN	9		NO_MNAR	Small	KNN	6
SE_MNAR	Small	HotDeck	0		NO_MNAR	Small	HotDeck	0
SE_MNAR	Small	Mode	0		NO_MNAR	Small	Mode	0
SE_MNAR	Medium	LogReg	125		NO_MNAR	Medium	LogReg	93
SE_MNAR	Medium	RF	86		NO_MNAR	Medium	RF	42
SE_MNAR	Medium	CART	81		NO_MNAR	Medium	CART	36
SE_MNAR	Medium	KNN	24		NO_MNAR	Medium	KNN	14
SE_MNAR	Medium	HotDeck	0		NO_MNAR	Medium	HotDeck	0
SE_MNAR	Medium	Mode	0		NO_MNAR	Medium	Mode	0
SE_MNAR	Large	LogReg	1094		NO_MNAR	Large	LogReg	655
SE_MNAR	Large	RF	934		NO_MNAR	Large	RF	471
SE_MNAR	Large	CART	927		NO_MNAR	Large	CART	482
SE_MNAR	Large	KNN	51		NO_MNAR	Large	KNN	55
SE_MNAR	Large	HotDeck	0		NO_MNAR	Large	HotDeck	0
SE_MNAR	Large	Mode	0		NO_MNAR	Large	Mode	0
SE_MNAR	Inclusive	LogReg	1516		NO_MNAR	Inclusive	LogReg	871
SE_MNAR	Inclusive	RF	1296		NO_MNAR	Inclusive	RF	624
SE_MNAR	Inclusive	CART	1311		NO_MNAR	Inclusive	CART	637
SE_MNAR	Inclusive	KNN	85		NO_MNAR	Inclusive	KNN	81
SE_MNAR	Inclusive	HotDeck	0		NO_MNAR	Inclusive	HotDeck	0
SE_MNAR	Inclusive	Mode	0		NO_MNAR	Inclusive	Mode	0

Note: *M=20 for all the procedures.*
Source: *compiled by the authors.*