Rethinking the Match: A Simulation-Based Assessment of Congeniality in continuous Prediction Models

Merlin Urbanski

Program: Methodology and Statistics for Behavioral, Biomedical, and Social Sciences
Supervisors: Dr. Maarten van Smeden (UMC Utrecht),
Dr. Anne de Hond (UMC Utrecht),
PhD Candidate Alex Carriero (UMC Utrecht)
Host Institution: Julius Center for Health Science and Primary Care, UMC Utrecht
Candidate Journal: Statistics in Medicine
FETC-approved: 24-2006 and 24-2242
Date: May 12, 2025
Wordcount: 5218

Abstract

Introduction

Missing data is a common challenge in medical research, and selecting an appropriate imputation method is crucial for accurate predictions. Congeniality refers to the alignment between the assumptions of the imputation model and the substantive prediction model. While this concept is well-understood in the context of parameter estimation, its implications for predictive performance and model calibration remain unclear.

Methods

We evaluated congenial and uncongenial model combinations across various scenarios, reflecting different relationships between predictors and the outcome. Our analysis focused on predictive accuracy and calibration in settings with continuous predictors and continuous outcomes. To illustrate these findings, we conducted a case study using the MIMIC-III dataset.

Results

Across all simulation scenarios, congeniality had no observable impact on the accuracy or calibration of model combinations. However, patterns from both the simulation study and the MIMIC-III case study suggested that interactions between the imputation model and the substantive prediction model can influence overall performance.

Conclusion

Accuracy and calibration are not determined by congeniality, while the combination of imputation and substantive prediction model matters.

1 Introduction

In many scientific disciplines, especially in medical research, missing data is a common and challenging issue. Patient information is often gathered from multiple hospitals, countries, and healthcare providers, each with distinct documentation practices and varying levels of detail. Moreover, not every patient receives the same diagnostic tests or treatments, resulting in datasets where measurements are not available for all patients across settings.

Traditional ways to deal with missing data in clinical prediction models include complete-case analysis (excluding incomplete cases) and mean-value imputation (filling missing values with averages). However, complete-case analysis tends to perform poorly in predictive settings since not all available information is used to train the prediction model [1]. Similarly, mean-value imputation only performs well under specific missing data mechanisms and when the mean carries meaningful information for the variable being imputed [2]. Although more advanced imputation methods that can handle more complex missing data situations are available, simple approaches like complete-case analysis and mean imputation remain common in practice, likely because they are easy to implement [3].

Imputation methods such as predictive mean matching (PMM), Bayesian regression imputation (B-RI), and random forest (RF) imputation have recently gained popularity for preparing data to develop prediction models, especially within the multiple imputation framework [4, 5]. At the same time, prediction algorithms such as random forest have also gained popularity as a new generation of prediction models that are able to model complex relationships [3]. Although several simulation studies have examined the predictive performance of different imputation models, they typically focus on a single substantive prediction model in these comparisons [1, 6].

Focusing on the performance of individual imputation methods in isolation overlooks an important aspect: the potential interaction between the imputation model and the substantive prediction model. Research suggests that a phenomenon called uncongeniality can have a strong influence on model performance, particularly in the context of parameter estimation. Uncongeniality was first described by Meng [7] as a scenario where the imputation model and the substantive model are not derived from a common joint distribution or do not share the same underlying assumptions. Existing research has shown that such uncongenial model combinations can lead to biased parameter estimates and invalid inferences [8, 4].

There is relatively little research on how uncongeniality affects performance in a predictive setting. As Orloagh ([9] p.321) puts it, the concepts of congeniality and uncongeniality were "developed for the setting of parameter estimation, and it is not clear that this matters in the prediction context." This highlights a gap in the current literature: while we know that uncongeniality can be problematic for inference, it remains unclear whether the same holds true for prediction, where the goal is not to estimate parameters accurately but rather to maximize predictive accuracy.

Based on this research gap, we evaluated the predictive performance of several congenial and uncongenial model combinations with continuous outcome. These included imputation methods from the `mice` package, combined with prediction models based on regression and random forests. We present the design of our simulation study in Section 2. In Section 3 we report the results. Section 4 describes the MIMIC-III case study and its findings. Section 5 discusses the implications of our simulation and case study results for researchers imputing values for prediction models. Finally, Section 6 offers our main conclusions.

2 Methods

The study adheres to the ADEMP (Aims, Data-generating mechanisms, Estimands, Methods, Performance measures) guidelines for the design and reporting of simulation studies [10]. All scripts and code used in this project can be accessed via GitHub: https://github.com/MerlinUrbanski/Uncongeniality.

2.1 Aim

This study investigates how congeniality between an imputation model and a substantive prediction model affects predictive performance. We compare model combinations across different univariate missingness scenarios, using only continuous variables and assess out-of-sample predictive performance.

2.2 Data-Generating Mechanisms

2.2.1 Scenarios

Data with continuous predictors and a continuous outcome will be simulated to reflect 40 ($5 \times 2 \times 2 \times 2$) unique scenarios. We vary four data characteristics across the scenarios: the missingness mechanism (see subsubsection 2.2.5), the type of relationship

among predictors, the strength of predictor correlation (see subsubsection 2.2.3), and the predictor–outcome relationship form (see subsubsection 2.2.4). The individual levels of each varying factor are shown in Table 1 and details of the data-generating processes will be explained in the following sections.

Table 1: Summary of factors to be varied in the simulation study. Data will be simulated to reflect 40 unique scenarios ($5 \times 2 \times 2 \times 2$), by varying the following characteristics. This table contains the following abbreviations: Missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR).

Factor	**Levels**
Missingness Mechanism	MCAR weak MAR, strong MAR weak MNAR, strong MNAR
Cor.-type between predictors	Linear, Quadratic
Cor.-type between predictors and outcome	Linear, Quadratic
Cor.-strength among predictors	Low, High

2.2.2 Generation of $X_2, \ldots, X_8$

For all scenarios we generate eight continous predictors $X_1, \ldots, X_8$ and one continous outcome variable Y. Seven of the eight predictors $X_2, \ldots, X_8$ will be generated from a multivariate normal distribution:

$$\mathbf{X}_{2:8} \sim \mathcal{N}_7(\mathbf{0},\, \mathbf{\Sigma}),$$

where the mean vector is given by

$$\mathbf{0} = (0, 0, \ldots, 0)^\top,$$

and the covariance matrix $\mathbf{\Sigma}$ is defined as

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix},$$

with $\rho \in \{0.2, 0.8\}$, where $\rho = 0.2$ represents the low correlation scenario and $\rho = 0.8$ represents the high correlation scenario.

2.2.3 Generation of X_1

Rather than sampling X_1 jointly with the other predictors, we construct it as either a linear or a quadratic function of $X_2, \ldots, X_8$ while approximately preserving the target correlation ρ. This construction facilitates a direct comparison of monotonic (linear) versus non-monotonic (quadratic) effects under identical correlation settings (something not achievable when all variables are drawn jointly from a multivariate normal distribution). A pitfall is that X_1 will generally have a different variance than the other predictors: in the low-correlation $\mathrm{Var}(X_1) \approx 0.8$, and in the high-correlation $\mathrm{Var}(X_1) \approx 1.5$. This may play a role when interpreting results, as variance differences may influence downstream model performance.

Linear Relationship with X_1

In the linear scenarios, we define X_1 as a weighted sum of the other predictors plus Gaussian noise:

$$X_{1i} = a \sum_{j=2}^{8} X_{ji} + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}(0,\, \sigma_e^2).$$

The coefficient a determines how strongly X_1 co-varies with $X_2, \ldots, X_8$, while the error variance σ_e^2 adds just enough noise to achieve our target correlation ρ. To find a and σ_e, we note that

$$\mathrm{Cov}(X_{1i}, X_{ji}) = a\,(1 + 6\rho), \quad \mathrm{Var}(X_{1i}) = a^2\,(7 + 42\rho) + \sigma_e^2,$$

and we solve

$$\mathrm{Corr}(X_{1i}, X_{ji}) = \frac{\mathrm{Cov}(X_{1i}, X_{ji})}{\sqrt{\mathrm{Var}(X_{1i})}} = \rho.$$

This yields the numerical values used in the simulations:

$$\begin{aligned} \rho = 0.2: &\quad a = \tfrac{1}{11}, \quad \sigma_e = \sqrt{0.6706}, \\ \rho = 0.8: &\quad a = \tfrac{0.8}{5.8}, \quad \sigma_e = \sqrt{0.228}. \end{aligned}$$

Quadratic Relationship with X_1

For a quadratic relationship, we square each predictor and add noise:

$$X_{1i} = a \sum_{j=2}^{8} X_{ji}^2 + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0,\, \sigma_e^2).$$

We use the same values of a and σ_e derived in the linear scenario, ensuring that the overall strength of relationships remains constant in both low- and high-correlation settings. It is important to be aware that, because the quadratic relationship is nonlinear and X_1 is no longer normally distributed, Pearson's correlation coefficient no longer provides a valid measure of association strength. When comparing the two scenarios with Pearson's correlation, the coefficient for the quadratic relationship will always be lower than in the linear case.

2.2.4 Generation of the Outcome Variable Y

When generating the outcome variable Y, we adopt a standard approach in which only a subset of the predictors directly affects Y [11]. Out of the 8 predictors, the first two are assumed to have a strong relationship, the next two a weak relationship, and the remaining four no direct relationship with the outcome (see Figure 2). To capture unexplained variability, we include an independent noise term $\varepsilon_i \sim \mathcal{N}(0, 1)$. The relationship with Y can be either linear or quadratic:

Linear Relationship with the Outcome

$$Y_i = 1.5\, X_{1i} + 1.5\, X_{2i} + 0.5\, X_{3i} + 0.5\, X_{4i} + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, 1).$$

Quadratic Relationship with the Outcome

$$Y_i = 1.5\, X_{1i}^2 + 1.5\, X_{2i}^2 + 0.5\, X_{3i}^2 + 0.5\, X_{4i}^2 + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, 1).$$

The variance of Y differs substantially between the linear and quadratic scenarios (see different scales in Figure 1). This difference in variance complicates fair comparison of variance-dependent performance metrics (e.g., Root Mean Squared Error (RMSE)), since models may appear to perform better or worse simply due to differences in outcome scale rather than true predictive ability. Therefore, metrics like RMSE should be used to compare models only within a single scenario and not between different scenarios. Figure 2 shows a schematic of all relationships that are being varied, and Figure 1 illustrates the outcome variable under each scenario.

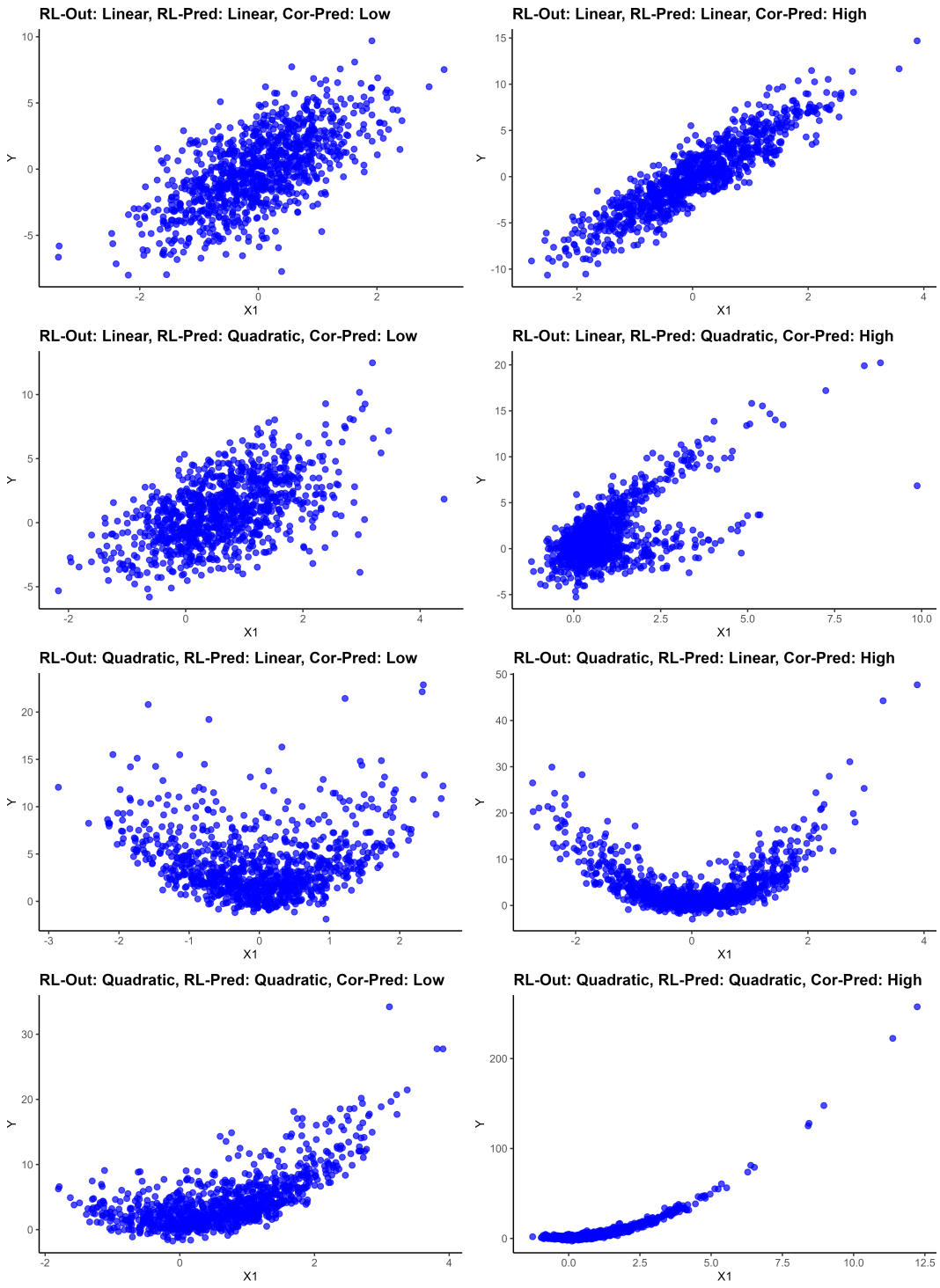


Figure 1: Visualization of the relationship between X1 and the outcome Y in eight ($2 \times 2 \times 2$) distinct scenarios. In the scenarios the type relationship between the predictors and outcome (RL-OUT), the type of relationship among predictors (RL-Pred) and strength of the relationship among predictors (Cor-Pred) varied. The five missing mechanisms are not visualised in this graph.

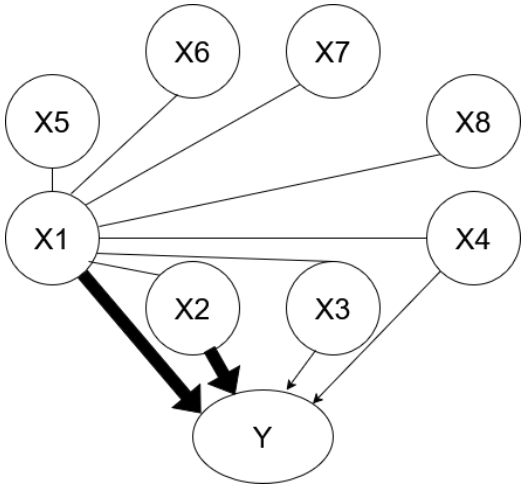


Figure 2: Visualization of the generated data. The lines among X1 and the other predictors visualise the relationships that can be linear or quadratic. The thick arrows on Y are the strong effects and the thin arrows are the weak effects on the outcome.

2.2.5 Missing Data Mechanisms

In the training data, missingness is imposed on the variable X_1 using five distinct mechanisms. For each mechanism, an indicator M is generated (with $M = 1$ denoting a missing value) by drawing from a Bernoulli distribution with a probability determined by the mechanism.

1. **MCAR (Missing Completely At Random):**

   Under the MCAR mechanism, every observation has the same probability of being missing, regardless of any other variable. Specifically, the probability of missingness is fixed at:

   $$P(M = 1) = 0.3.$$

2. **Weak MAR (Missing At Random):**

   In the weak MAR scenario, the probability that an observation is missing depends partly on the covariate X_2 and partly on random variation. First, the rank of X_2 is calculated. Then, the missingness probability is defined as a weighted combination of the rank-based component and a random component:

   $$P(M = 1) = w \cdot \frac{\text{rank}(X_2)}{n} + (1 - w) \cdot U(0, 1),$$

where:

- w is a weight (set to 0.5) that controls the relative influence of the rank-based term,
- n is the total number of observations,
- $U(0,1)$ denotes a random value drawn from a uniform distribution on the interval $[0,1]$.

3. **Strong MAR:**

   For the strong MAR mechanism, missingness is driven solely by X_2 without added noise:

$$P(M=1) = \frac{\text{rank}(X_2)}{n},$$

4. **Weak MNAR (Missing Not At Random):**

   In the weak MNAR scenario the probability of being missing depends on X_1 itself and added random noise (w is set to 0.5):

$$P(M=1) = w \cdot \frac{\text{rank}(X_1)}{n} + (1-w) \cdot U(0,1),$$

5. **Strong MNAR:**

   Under the strong MNAR mechanism, the missingness is determined exclusively by X_1:

$$P(M=1) = \frac{\text{rank}(X_1)}{n}$$

In each case, once the probability is computed, it is rescaled so that its average equals the intended missing percentage of 0.3. If the provisional probability for an observation is denoted by p_i and the mean of these provisional probabilities across all observations is $\bar{p}$, then each probability is replaced by

$$p_i^* = p_i \times \frac{0.3}{\bar{p}},$$

which guarantees that the average of $\{p_i^*\}$ is close to 0.3. The final missingness indicator for each observation is then drawn from a Bernoulli distribution with success probability p_i^*. The 5 different missing mechanisms are visualised in Figure 3.

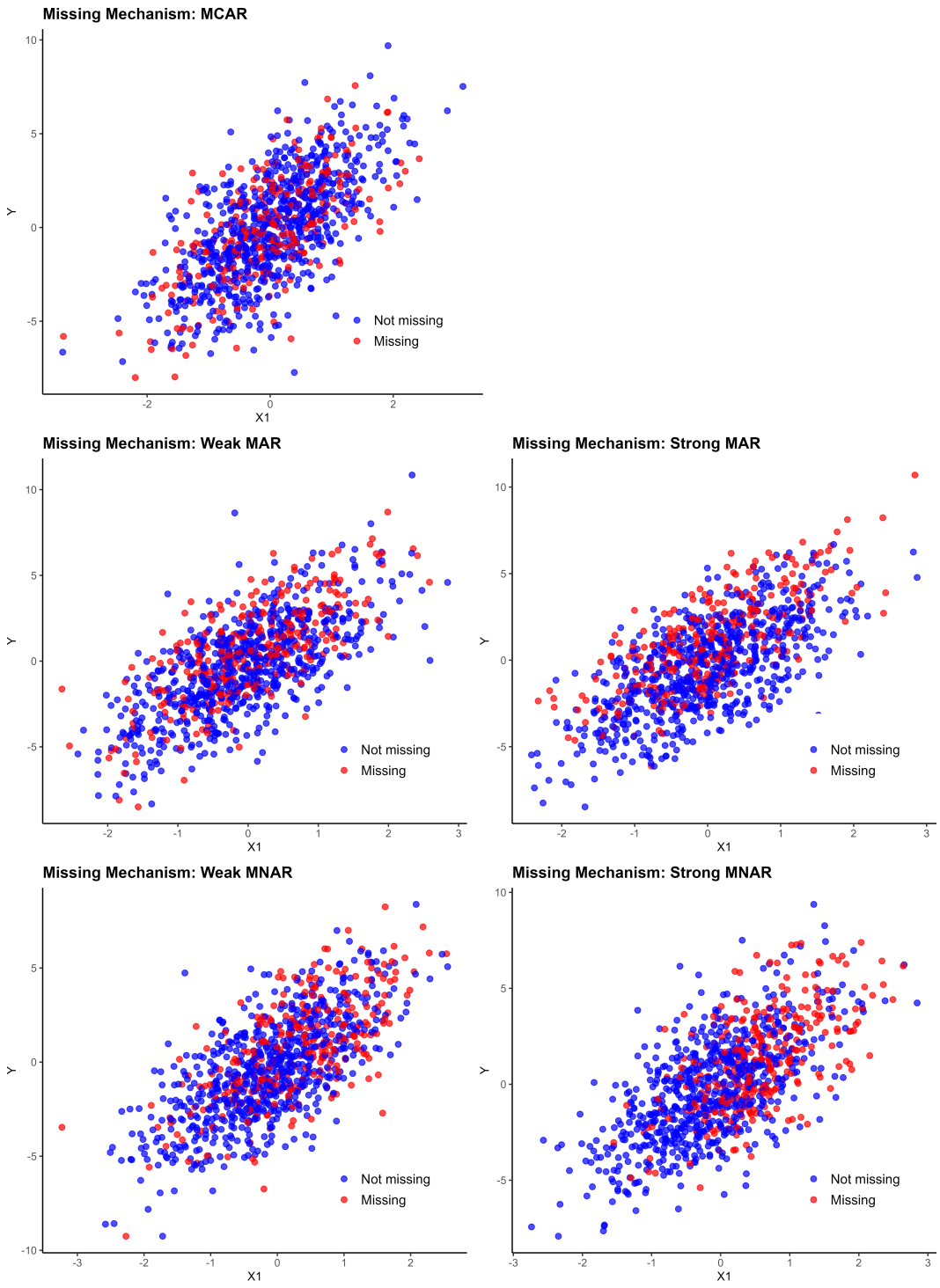


Figure 3: Visualization of the 5 Missing Mechanisms (missing completely at random (MCAR), weak missing at random (weak MAR), strong missing at random (strong MAR), weak missing not at random (weak MNAR), strong missing not at random (strong MNAR)) in a low correlation scenario with only linear relationships.

2.3 Estimands

We evaluated the predictive performance of the model combinations using a large out-of-sample validation set (see subsection 2.5).

2.4 Methods

2.4.1 Imputation Methods

To address the univariate missing data, we used imputation methods implemented in the widely used R package `mice` [12]. We evaluated five imputation strategies (see also Table 2): (1) Predictive Mean Matching (PMM) (`mice::pmm`), (2) PMM extended to include quadratic terms (`mice::quadratic`), (3) regression imputation (`mice::norm.predict`), (4) regression imputation with manually specified squared terms, and (5) random forest imputation (`mice::rf`).

To save computational time and because our primary goal was to obtain the most accurate imputed values for prediction (rather than to quantify imputation uncertainty for inference) we opted for a single imputation approach instead of multiple imputation. Single imputation requires deterministic algorithms [13], so we limited predictive mean matching to a single donor and replaced Bayesian regression imputation with its deterministic regression imputation. Random forest (RF) imputation necessarily remains stochastic, so we relied on the default RF implementation in `mice`. For all imputation methods, the number of iterations was kept at the default value of 5.

Table 2: Imputation Strategies Employed

Full Name	Abbreviation	R/`mice` Command
Predictive Mean Matching	PMM	`mice::pmm`
PMM with Quadratic Terms	PMM-Q	`mice::quadratic`
Regression Imputation	RI	`mice::norm.predict`
Regression Imputation with Quadratic Terms	RI-Q	`mice::norm.predict` (with quadratic terms)
Random Forest Imputation	RF	`mice::rf`

2.4.2 Prediction Models

In our study, we employed two distinct predictive modeling approaches (see Table 3), both of which were trained on the imputed training dataset and subsequently applied to the validation data (see subsubsection 2.4.4). The first approach was a regression model that incorporated both linear and quadratic terms for all predictors enabling to capture potential linear and non-linear effects.

The second approach involved a random forest algorithm. Recognizing the tendency of tree-based methods to overfit, we opted to use 5-fold cross-validation for hyperparameter tuning [14]. On each individual training data set, we conducted a comprehensive grid search across several hyperparameters, specifically varying the number of trees (500, 1000, and 1500), the nodesize (2, 5, and 8), and the number of predictors sampled at each split (3, 5, and 7). The optimal model, determined by the cross-validation results, was then applied to the validation data.

Table 3: Employed Prediction Models

Model Type	Abbreviation	R Package
Regression Model (Linear & Quadratic)	Reg	base R (`lm`)
Random Forest	RF	`randomForest`

2.4.3 Model Combinations and Congeniality

The five imputation models and the two substantive prediction models result in ten (5×2) distinct model combinations (see Table 4). We considered six model combinations uncongenial. Five of these combinations differ in their assumptions regarding the complexity of relationships and occur when a random forest model (which is capable of modeling complex relationships)is paired with either regression-based or PMM, methods that assume simpler relationships. The final uncongenial model combination (RI + REG) is classified as such because the regression imputation model assumes only linear effects, whereas the regression prediction model can also capture quadratic effects.

Two model combinations were labeled "Rather Uncongenial" and "Rather Congenial." Both of these include a PMM-based imputation method (PMM and PMM-Q) and a regression prediction model (REG). We chose not to classify them strictly as uncongenial or congenial because PMM is often described as a semi-parametric

method, viewed as a hybrid of parametric regression and the non-parametric k-nearest neighbor approach [15]. This results in PMM having less strong assumptions than those of fully parametric regression, so the differences in assumptions are less pronounced. Finally, there are two congenial model combinations (RI-Q + REG and RF + RF) in which the imputation model is essentially identical to the prediction model, sharing the same underlying assumptions.

Table 4: Congeniality of Model Combinations.

Imputation Model	**Prediction Model**	**Congeniality**
Predictive mean matching	Linear regression	Rather uncongenial
Predictive mean matching (Quadratic)	Linear regression	Rather congenial
Regression imputation	Linear regression	Uncongenial
Regression imputation (Quadratic)	Linear regression	Congenial
Random forest imputation	Linear regression	Uncongenial
Predictive mean matching	Random forest	Uncongenial
Predictive mean matching (Quadratic)	Random forest	Uncongenial
Regression imputation	Random forest	Uncongenial
Regression imputation (Quadratic)	Random forest	Uncongenial
Random forest imputation	Random forest	Congenial

2.4.4 Simulation Methods

For each simulation scenario, 100 data sets were generated. Each data set consisted of a training set and a corresponding validation set, both drawn independently using the same data-generating mechanism. Missing values were introduced only in the training data; the validation data were generated without any missingness. To enable precise performance evaluation, validation sets were made substantially larger than training sets, with 1,000 observations in the training data and 100,000 in the validation data.

As described in subsection 2.4, ten model combinations (5×2) were developed for each simulated data set. Model training was performed on the training data, and predictive performance was assessed using the validation data.

2.5 Performance Measures

The performance measures in this study are metrics that evaluate the relationship between the predicted values ($\hat{Y}$) and the observed outcome values (Y) in the validation data.

1. **Root Mean Squared Error (RMSE):** RMSE evaluates the average size of error between the predicted and true values, providing a measure of overall predictive accuracy. It is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}.$$

   Lower RMSE values indicate more accurate predictions.

2. **Coefficient of Determination (R^2):** R^2 measures the proportion of variance in the true outcomes (Y) that is explained by the predicted values ($\hat{Y}$). It is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}.$$

   R^2 ranges from 0 to 1, where $R^2 = 0$ indicates that the model explains none of the variability in Y, and $R^2 = 1$ indicates it explains all the variability perfectly.

3. **Evaluation of flexible Calibration Curves:** Calibration is assessed through visual inspection of calibration plots, which visualize the relationship between predicted and observed values. Deviations from the ideal 45 ° 1:1 line indicate systematic over- or underestimation, providing insights into the alignment of predictions with observed outcomes. Calibration curve coordinates were estimated using LOESS regression, with 200 points extracted from the fitted curve for plotting. The resulting curves were visualized with `ggplot2` [16].

2.6 Software

Simulations and result processing were conducted in R (v4.4.0) [17] using the High Performance Computer at the University Medical Center Utrecht.

3 Results

Model performance varied substantially across the simulation scenarios. However, missing data mechanisms had little influence on RMSE and R^2 across all scenarios (see Appendix A). Accordingly, we focused our result section on scenarios that differed in the type of the relationship between predictors and the outcome, the type of relationships among the predictors and their strengths. Furthermore, in the following sections we will focus mainly on the RMSE and not the R^2 since they are strongly related. The R^2 values are still shown in the tables but not mentioned in the text. It is important to note that the variance of the outcome variable Y differs strongly due to the different data generating mechanisms. Therefore, it is not advised to directly compare the RMSE between scenarios. Instead we compared the ranking of the different model combinations between scenarios.

Similarly to RMSE and R^2, the missing data mechanisms had a minimal impact on the calibration plots (see Appendix B). We therefore compare only the missing completely at random (MCAR) scenarios in this result section.

3.1 RMSE and R^2 in different scenarios

3.1.1 Linear Relationships among Predictors and between Predictors and Outcome

When considering combinations with a regression (REG) prediction model, those using PMM-based imputation models (PMM and PMM-Q) performed best across both high- and low-correlation scenarios (see Table 5). The model combination including RF-imputation performed better in the high correlation scenario but overall the differences between model combinations fall within one standard deviation.

When examining combinations using a RF prediction model, those using PMM as the imputation method performed best, followed by combinations based on regression imputation (RI and RI-Q) in both high- and low-correlation scenarios. Model combinations using PMM-Q consistently performed worse than other combinations, despite PMM performing well. Model combinations that were uncongenial or rather congenial/uncongenial outperformed the congenial model combinations.

Table 5: Mean RMSE and mean R^2 of model combinations under scenarios with linear relationships among predictors and linear relationships between predictors and outcome. Based on the RMSE the three best-performing combinations are highlighted from darkest (best) to lightest; the two combinations with the highest RMSE are unhighlighted. Abbreviations for model combinations: Predictive mean matching (PMM), predictive mean matching with quadratic effects (PMM-Q), regression imputation (RI), regression imputation with quadratic effects (RI-Q), random forest (RF), regression prediction (REG).

Imputation Model	**Prediction Model**	**Low Correlation**				**High Correlation**				**Congeniality**
		RMSE		R^2		**RMSE**		R^2		
		Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	
PMM	REG	1.013	0.024	0.868	0.008	1.011	0.024	0.931	0.004	Rather Uncongenial
PMM-Q	REG	1.013	0.024	0.868	0.008	1.010	0.024	0.931	0.004	Rather Congenial
RI	REG	1.029	0.026	0.865	0.008	1.031	0.026	0.929	0.004	Uncongenial
RI-Q	REG	1.031	0.027	0.865	0.008	1.033	0.026	0.928	0.004	Congenial
RF	REG	1.033	0.027	0.864	0.008	1.019	0.024	0.930	0.004	Uncongenial
PMM	RF	1.122	0.029	0.840	0.010	1.076	0.026	0.922	0.005	Uncongenial
PMM-Q	RF	1.141	0.035	0.834	0.011	1.095	0.034	0.919	0.006	Uncongenial
RI	RF	1.125	0.030	0.837	0.010	1.084	0.027	0.921	0.005	Uncongenial
RI-Q	RF	1.125	0.030	0.837	0.010	1.084	0.027	0.921	0.005	Uncongenial
RF	RF	1.159	0.032	0.833	0.010	1.087	0.027	0.921	0.005	Congenial

3.1.2 Quadratic Relationships among Predictors and Linear Relationships between Predictors and Outcome

For combinations involving a regression prediction model, the rankings in low-correlation scenarios closely mirrored those observed in earlier purely linear settings (see Table 6). Both PMM-based methods performed best, followed closely by RI-based methods. Methods including quadratic effects (PMM-Q, RI-Q) performed slightly worse than their linear-only counterparts (PMM, RI).

In high-correlation scenarios, the performance ranking shifted. Regression-based imputation methods (RI, RI-Q) performed similarly to PMM, followed by PMM-Q and RF methods. This is the only scenario where a congenial model combination performed the best. In the three ramaining scenarios of Table 6, the congenial combinations were outperformed by uncongenial and by rather congenial/uncongenial combinations.

For scenarios using RF prediction models, performance rankings remained consistent across both low and high-correlation scenarios. Model combinations using RI-based imputations consistently outperformed PMM-based imputations. Additionally, imputation methods incorporating quadratic effects consistently performed worse than those using linear effects alone.

Table 6: Mean RMSE and mean R^2 of model combinations under scenarios with quadratic relationships among predictors and linear relationships between predictors and outcome. Based on the RMSE the three best-performing combinations are highlighted from darkest (best) to lightest; the two combinations with the highest RMSE are unhighlighted. Abbreviations for model combinations: Predictive mean matching (PMM), predictive mean matching with quadratic effects (PMM-Q), regression imputation (RI), regression imputation with quadratic effects (RI-Q), random forest (RF), regression prediction (REG).

Imputation Model	**Prediction Model**	**Low Correlation**				**High Correlation**				**Congeniality**
		RMSE		R^2		**RMSE**		R^2		
		Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	
PMM	REG	1.016	0.025	0.836	0.010	1.031	0.026	0.893	0.010	Rather Uncongenial
PMM-Q	REG	1.018	0.026	0.835	0.011	1.041	0.034	0.891	0.011	Rather Congenial
RI	REG	1.027	0.027	0.835	0.011	1.031	0.030	0.895	0.010	Uncongenial
RI-Q	REG	1.030	0.027	0.834	0.011	1.030	0.025	0.893	0.010	Congenial
RF	REG	1.040	0.029	0.830	0.011	1.055	0.034	0.888	0.011	Uncongenial
PMM	RF	1.130	0.030	0.800	0.013	1.117	0.039	0.875	0.011	Uncongenial
PMM-Q	RF	1.155	0.039	0.791	0.016	1.144	0.048	0.869	0.013	Uncongenial
RI	RF	1.124	0.029	0.800	0.013	1.105	0.040	0.877	0.011	Uncongenial
RI-Q	RF	1.125	0.029	0.800	0.013	1.112	0.038	0.876	0.011	Uncongenial
RF	RF	1.174	0.033	0.789	0.014	1.142	0.040	0.870	0.011	Congenial

3.1.3 Linear Relationships among Predictors and Quadratic Relationships between Predictors and Outcome

The ranking of model combinations differed compared to the previous scenarios. When looking at combinations with a regression prediction model the ranking of model combinations in low and high correlation scenarios appears similar (see Table 7). In both scenarios combinations including PMM-Q and RI are performing almost equally well, outperforming the other model combinations. Model combinations including PMM, perform far worse compared to the other model combinations.

For RF prediction models, the ranking of model combinations varies considerably between low and high correlation scenarios. In the low correlation scenario, PMM-Q performs best, followed by RF and RI-Q. In contrast, under high correlation, RI ranks highest, with RI-Q and PMM-Q following. As with regression models, model combinations using PMM consistently perform the worst.

Within model combinations using a REG prediction model, the congenial model combinations consistently ranked third out of five. For model combinations using a RF prediction model, the congenial model combinations performed second best in the low correlation scenario, but dropped to fourth out of five in the high correlation scenario.

Table 7: Mean RMSE and mean R^2 of model combinations under scenarios with linear relationships among predictors and quadratic relationships between predictors and outcome. Based on the RMSE the three best-performing combinations are highlighted from darkest (best) to lightest; the two combinations with the highest RMSE are unhighlighted. Abbreviations for model combinations: Predictive mean matching (PMM), predictive mean matching with quadratic effects (PMM-Q), regression imputation (RI), regression imputation with quadratic effects (RI-Q), random forest (RF), regression prediction (REG).

Imputation Model	**Prediction Model**	**Low Correlation**				**High Correlation**				**Congeniality**
		RMSE		R^2		**RMSE**		R^2		
		Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	
PMM	REG	1.138	0.051	0.876	0.014	1.115	0.044	0.950	0.006	Rather Uncongenial
PMM-Q	REG	1.073	0.038	0.894	0.011	1.033	0.027	0.958	0.005	Rather Congenial
RI	REG	1.073	0.032	0.894	0.011	1.030	0.030	0.958	0.005	Uncongenial
RI-Q	REG	1.076	0.037	0.893	0.011	1.052	0.049	0.956	0.006	Congenial
RF	REG	1.078	0.039	0.890	0.011	1.072	0.037	0.954	0.006	Uncongenial
PMM	RF	1.579	0.099	0.769	0.023	1.519	0.164	0.915	0.013	Uncongenial
PMM-Q	RF	1.487	0.097	0.793	0.018	1.460	0.175	0.922	0.014	Uncongenial
RI	RF	1.513	0.096	0.787	0.020	1.435	0.166	0.923	0.013	Uncongenial
RI-Q	RF	1.511	0.096	0.787	0.020	1.458	0.169	0.921	0.014	Uncongenial
RF	RF	1.508	0.098	0.789	0.019	1.477	0.165	0.920	0.013	Congenial

3.1.4 Quadratic Relationships among Predictors and between Predictors and Outcome

Combinations with a regression prediction model performed best with PMM based imputation methods in high and low correlation scenarios with PMM-Q outperforming PMM (see Table 8). The difference among combinations is relatively small in the low correlation scenarios but there are big difference in the high correlation scneario between the three most accurate model combinations and the two least accurate combinations (RI + REG and RF + REG).

For models with a RF prediction model the ranking of combinations differed. In the low correlation scenario RI-Q performed the best followed by model combinations including PMM based methods (PMM and PMM-Q). In the high correlation scenario the combination including PMM performed best followed by the REG based imputation methods (RI and RI-Q). The difference between combinations was not as high as in combinations with a REG prediction model but the standard deviation was extremely high.

Among model combinations using a REG prediction model, the congenial consistently ranked third across both low and high correlation scenarios. In contrast, when using RF as the prediction model, the congenial combination was consistently outperformed by uncongenial combinations.

Table 8: Mean RMSE and mean R^2 of model combinations under scenarios with quadratic relationships among predictors and quadratic relationships between predictors and outcome. Based on the RMSE the three best-performing combinations are highlighted from darkest (best) to lightest; the two combinations with the highest RMSE are unhighlighted. Abbreviations for model combinations: Predictive mean matching (PMM), predictive mean matching with quadratic effects (PMM-Q), regression imputation (RI), regression imputation with quadratic effects (RI-Q), random forest (RF), regression prediction (REG).

Imputation Model	**Prediction Model**	**Low Correlation**				**High Correlation**				**Congeniality**
		RMSE		R^2		**RMSE**		R^2		
		Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	
PMM	REG	1.108	0.054	0.930	0.009	1.479	0.603	0.987	0.014	Rather Uncongenial
PMM-Q	REG	1.103	0.052	0.930	0.009	1.332	0.439	0.988	0.011	Rather Congenial
RI	REG	1.167	0.090	0.924	0.013	2.585	1.270	0.959	0.032	Uncongenial
RI-Q	REG	1.127	0.115	0.926	0.018	1.575	0.768	0.982	0.022	Congenial
RF	REG	1.130	0.073	0.926	0.012	2.330	0.854	0.964	0.025	Uncongenial
PMM	RF	1.680	0.139	0.837	0.019	3.381	1.865	0.930	0.061	Uncongenial
PMM-Q	RF	1.678	0.140	0.838	0.019	3.557	1.970	0.924	0.066	Uncongenial
RI	RF	1.727	0.139	0.827	0.020	3.434	1.814	0.930	0.058	Uncongenial
RI-Q	RF	1.618	0.136	0.847	0.018	3.384	1.901	0.934	0.060	Uncongenial
RF	RF	1.693	0.162	0.839	0.023	3.857	1.882	0.919	0.060	Congenial

3.1.5 Performance across all scenarios

Analysis of the RMSE and R^2 revealed that model combinations incorporating a regression prediction model performed consistently better than those employing a random forest prediction model. In all scenarios it was never the case that within a scenario a combination using a regression prediction model performed better than a combination using a RF prediction model.

Among the combinations that used a regression prediction model, the approaches employing PMM-based imputation methods achieved the best performance followed by the combinations including regression based imputation models (see Table 9). In contrast, for combinations using a RF prediction model, regression-based imputation models outperformed the PMM-based combinations. Combinations using a RF imputation model constantly performed poorly in most scenarios compared to the other combinations.

Congeniality

The performance of model combinations varied significantly across different scenarios. No single model combination consistently ranked within the top two across all scenarios. Additionally, the effectiveness of imputation methods depended on the prediction model that it is combined as shown in Table 9. However, the congeniality or uncongeniality of the model combination did not determine whether a combination would perform well or poorly.

Table 9: Mean rank of RMSE and mean rank of R^2 by model combination across all scenarios. Based on the RMSE-Rank, the three best-performing combinations are highlighted from darkest (best) to lightest; the two combinations with the highest RMSE-Rank are unhighlighted. Abbreviations for model combinations: Predictive mean matching (PMM), predictive mean matching with quadratic effects (PMM-Q), regression imputation (RI), regression imputation with quadratic effects (RI-Q), random forest (RF), regression prediction (REG).

Imputation Model	Prediction Model	Mean Rank RMSE	SD	Mean Rank R^2	SD	Congeniality
PMM	REG	2.730	1.478	2.813	1.466	RU
PMM-Q	REG	2.122	1.163	2.211	1.135	RC
RI	REG	3.182	1.275	2.855	1.377	U
RI-Q	REG	3.217	1.374	3.184	1.366	C
RF	REG	3.749	1.210	3.937	1.101	U
PMM	RF	3.038	1.414	2.952	1.451	U
PMM-Q	RF	3.121	1.457	3.131	1.488	U
RI	RF	2.568	1.316	2.748	1.354	U
RI-Q	RF	2.400	1.143	2.541	1.189	U
RF	RF	3.872	1.232	3.628	1.322	C

3.2 Calibration Plots

Across all scenarios, calibration curves of combinations using Random Forest (RF) prediction models were shorter, covering a narrower range compared to those using regression (REG) models, which spanned a wider range of predicted values (see Figure 4 and Figure 5).

In scenarios involving quadratic effects either among predictors or between predictors and outcome, calibration curves of combinations with RF prediction models showed a notable spread at higher predicted values. The variation indicates that some combinations produced substantial underestimation in the upper range of predicted values, while others remained better calibrated in that region. This spread became more pronounced under conditions of high predictor correlation. Furthermore, these calibration curves consistently lay above the 45° line, indicating systematic underestimation of high values by RF models in quadratic scenarios. A similar pattern emerged in high-correlation scenarios with quadratic relationships among predictors when RF models were used for imputation.

Another type of spread occurred at high predicted values in scenarios with quadratic relationships among predictors but linear relationships with the outcome, particularly when using the RI + REG combination. This spread was more pronounced under high correlation conditions, with calibration lines mostly falling below the 45° line, indicating systematic overestimation. All other model combinations showed relatively good calibration with only minor deviations from the 45° line. Finally, congeniality did not affect the calibration of combinations.

Figure 4: Calibration Plots of all model combinations across different scenarios with a linear outcome-relationship (only missing completely at random (MCAR) scenarios). Abbreviations for scenario description: Relationship with outcome (RL-Out), relationship with predictors (RL-Pred), correlation among predictors (Cor-Pred). Abbreviations for model combinations: Predictive mean matching (PMM), predictive mean matching with quadratic effects (PMM-Q), regression imputation (RI), regression imputation with quadratic effects (RI-Q), random forest (RF), regression prediction (REG).

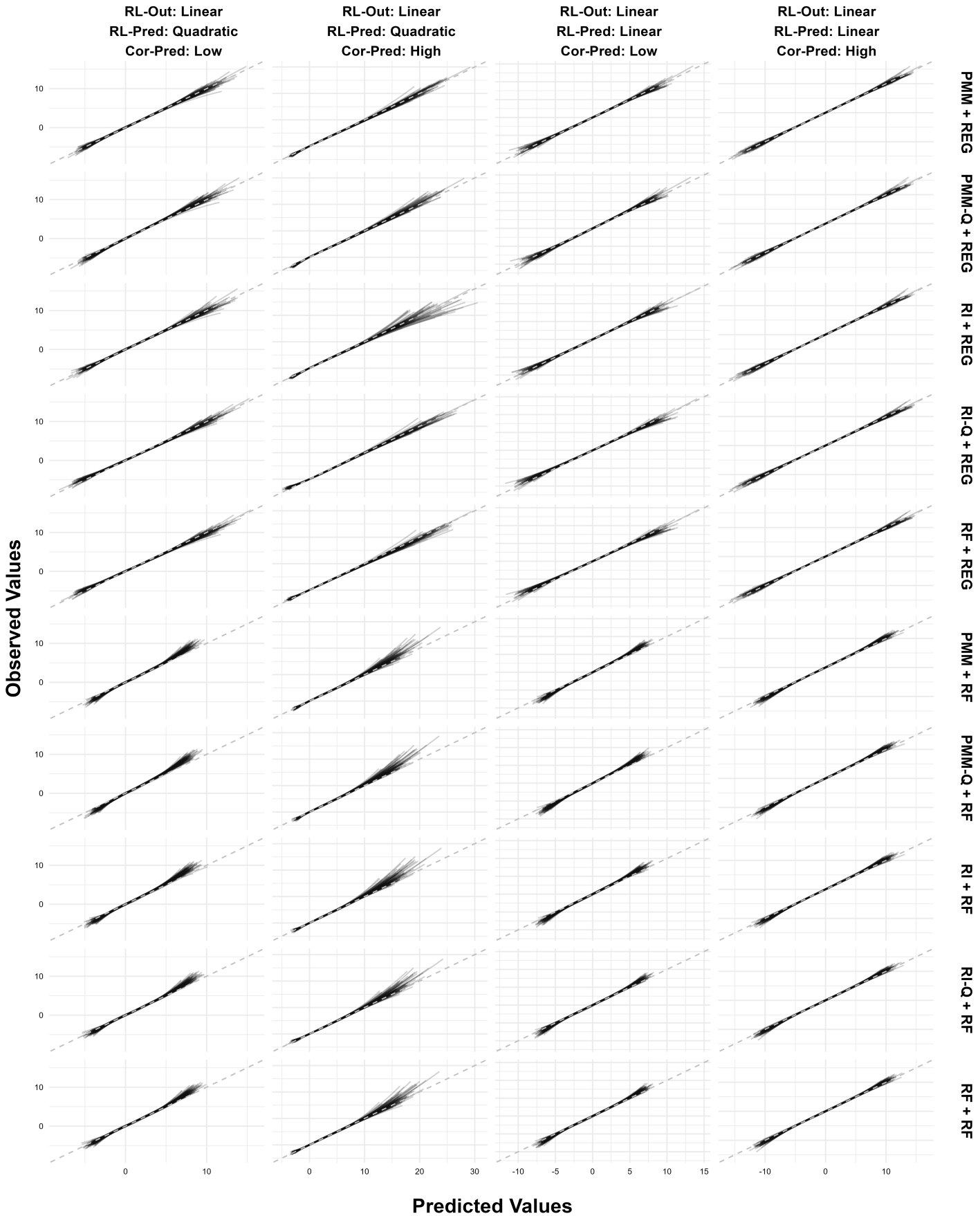
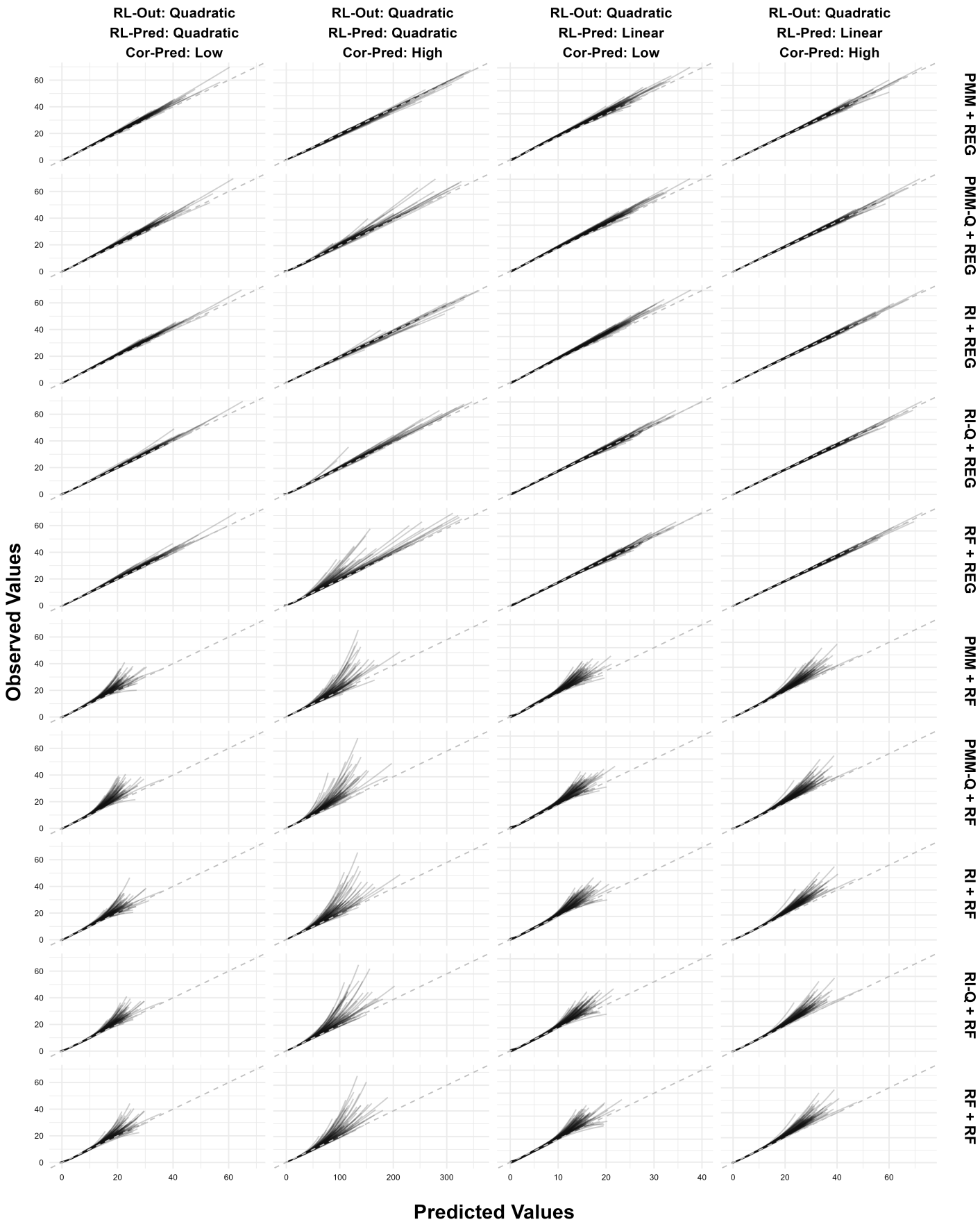


Figure 5: Calibration Plots of all model combinations across different scenarios with a quadratic outcome-relationship (only missing completely at random (MCAR) scenarios). Abbreviations for scenario description: Relationship with outcome (RL-Out), relationship with predictors (RL-Pred), correlation among predictors (Cor-Pred). Abbreviations for model combinations: Predictive mean matching (PMM), predictive mean matching with quadratic effects (PMM-Q), regression imputation (RI), regression imputation with quadratic effects (RI-Q), random forest (RF), regression prediction (REG).



4 MIMIC-III Data Case Study

To evaluate how our simulation findings translate to real-world data, we conducted a case study using the publicly available MIMIC-III database [18, 19]. We predicted each patient's mean blood urea nitrogen concentration (BUN) using the mean of nine routine laboratory tests: hemoglobin, hematocrit, white blood cell count, creatinine, glucose, sodium, potassium, chloride and bicarbonate. We trained and evaluated the same ten model combinations (5×2) from our simulation study.

The MIMIC-III cohort comprises data of 46,467 individual patients at Beth Israel Deaconess Medical Center between 2008 and 2014 [18, 19]. For each patient we computed the mean values of all variables across every measurement during their stay and we used those mean values to train and validate the model combinations. We excluded patients with missing mean values in any variable in order to later recreate a univariate missingness scenario, like in the simulation study.

Because our regression prediction model is very simple, it is especially vulnerable to extreme values. Since our dataset comes primarily from an intensive care unit, it likely contains extreme outliers in at least one of the ten biomarkers (for example, from intoxications or analytical errors). Such outliers can distort the model's estimates and lead to uninterpretable results. To prevent this, we excluded any patient who had one biomarker value more than four standard deviations from the mean. After removing cases with missing data and extreme outliers, 37,086 patients remained for analysis.

To match our simulation design, we divided the data into a training set containing 1,000 observations and a validation set containing the remaining 36,086 observations. In the training set, we then applied 30% missingness to Creatine (the predictor most strongly correlated with BUN) using the same MCAR procedure as in our simulation study (see subsubsection 2.2.5). The ten model combinations were then trained on the training set and applied to the validation set like in the simulation study (see subsection 2.4).

In Table 10, we present the RMSE and R^2 for each model combination. On the MIMIC dataset, random forest prediction models achieve slightly lower RMSE and higher R^2 than regression-based models. They also are better calibrated, with their calibration curves closely following the 45° line (see Appendix C). In contrast, the regression models tend to underestimate low BUN values and overestimate high values.

Like in the simulation study, congeniality did not influence the RMSE, R^2 or calibration curves of the model combinations. Instead the same pattern was observed. Regression prediction models achieve the lowest RMSE and highest R^2 with PMM-based imputation methods while random forest prediction models perform best with regression-based imputation methods.

Table 10: The RMSE and R^2 values of each model combination in the MIMIC-III data example. Abbreviations for model combinations: Predictive mean matching (PMM), predictive mean matching with quadratic effects (PMM-Q), regression imputation (RI), regression imputation with quadratic effects (RI-Q), random forest (RF), regression prediction (REG).

Imputation Model	**Prediction Model**	**RMSE**	R^2	**Congeniality**
PMM	REG	8.542	0.657	Rather Uncongenial
PMM-Q	REG	8.508	0.660	Rather Congenial
RI	REG	8.577	0.657	Uncongenial
RI-Q	REG	8.599	0.656	Congenial
RF	REG	8.653	0.648	Uncongenial
PMM	RF	8.489	0.667	Uncongenial
PMM-Q	RF	8.481	0.663	Uncongenial
RI	RF	8.428	0.666	Uncongenial
RI-Q	RF	8.444	0.665	Uncongenial
RF	RF	8.580	0.657	Congenial

5 Discussion

In this study, we investigated how the congeniality of model combinations affects the out-of-sample performance of clinical prediction models for a continuous outcome. Our results show that congeniality did not influence accuracy, R^2 or calibration. Instead, the ranking of combinations was influenced primarily by the type and strength of relationships among predictors and between predictors and the outcome. No single imputation model consistently outperformed the others across both substantive pre-

diction models (see Table 9), indicating that the pairing of imputation and prediction model matters. Although some prediction models perform better with certain imputation methods, this practical "compatibility" is unrelated to the formal definitions of compatibility or congeniality in the inferential-statistics literature.

Across all scenarios (see Table 9), two clear patterns emerged, despite the small differences between model combinations. Linear regression (REG) prediction models achieved their highest accuracy when paired with PMM-based imputation (PMM, PMM-Q), whereas random forest (RF) prediction models performed better following regression-based imputation (RI and RI-Q). This difference likely reflects how each algorithm uses the imputed data. Single-donor PMM often assigns identical values to multiple missing cases, reducing the variability that regression trees need to distinguish patterns. Regression imputation, by contrast, produces different values that preserve variability and thus may better support RF learning. The REG prediction model might handle identical imputed values more easily because they fit a single regression line rather than partitioning data across multiple regression trees. In fact, REG models consistently perform better with PMM-based imputation than with RI-based methods. However, the exact statistical reasons behind this apparent compatibility should be explored in future research.

That the congenial pairing of random forest imputation with random forest prediction (RF + RF) did not outperform the uncongenial combinations in either our simulation study or the MIMIC-III example may be due to multiple factors. As noted earlier, random forest prediction models may struggle with imputation methods that group observations (random forest is also an algorithm that groups observations like PMM). However, its low performance could also be because we left RF imputation at its default settings rather than tuning its hyperparameters. Applying the same tuning procedure to the RF imputation algorithm would probably improve the performance of the model combination. Furthermore, tuning both random forest models similarly would lead to a more genuine form of congeniality of the combination.

The poor performance of the other congenial combination (RI-Q + REG) may stem from regression prediction models struggling with regression-based imputation methods. An additional factor could be that in our simulation the RI-Q imputation model was not ideally specified for the true data-generating process. Specifically, RI-Q imputes using the squared outcome (Y), whereas when the relationship is

$$X_{1i} = \sqrt{\tfrac{2}{3}\, Y_i + \ldots}$$

it would be more appropriate to impute on the square root of the outcome. The `mice::quadratic` (PMM-Q) method addresses this by defining the outcome transformation internally, thus avoiding the mismatch.

Across all scenarios in our simulation study, RF prediction models were less accurate and had shorter calibration lines than REG prediction models. Furthermore, the calibration lines of RF models spread when there were quadratic relationships in the data, especially between predictors and the outcome. This may be because random forests struggle more in areas with few training data points, such as the tails of the distribution. Also, random forests do not extrapolate, meaning they can only predict within the range seen during training. If a new case falls outside that range, the model returns the closest value it learned and cannot capture truly extreme outcomes. Those difficult sparse regions are especially common under the quadratic data-generating mechanisms used in our simulations. In the applied MIMIC-III example, the RF prediction model performed better than the regression models. This suggests that our simulation's data-generating process likely favors regression prediction models, and that using more complex mechanisms would make the comparison between RF and REG prediction models fairer [20].

Another limitation is that the percentage of missing data is at only 30%, and missingness occurred in only one of the eight predictors. Although the variable with missing values (X_1) had a strong relationship with the outcome, it could be imputed effectively using the other seven predictors and the outcome itself. In other comparable studies, the overall percentage of missingness is often higher, or missing values are spread across multiple variables [21, 1, 8]. Furthermore, in our simulation study missingness was introduced only in the training set, while the validation (and deployment) data remained fully observed. In real-world applications, missing values often occur at validation or deployment, and it is advisable to use different strategies in those cases [13]. Consequently, our conclusions apply specifically to scenarios with univariate missingness introduced to the training data.

Finally, we relied on single imputation rather than the multiple-imputation framework, which is the default in `mice` [12]. While multiple imputation has been shown to improve the accuracy of binary regression predictions [1], we omitted it for practical reasons. First, hyperparameter tuning of a random forest on each imputed dataset would dramatically increase computational demands. Second, no established method exists for pooling random forest models across multiple imputations. We would need to train separate forests on each dataset and then average their predictions, an approach that is both uncommon in applied research and would introduce substantial

complexity. However, the random forest imputation method may profit from multiple imputation, since it is unlike the other imputation methods probabilistic by nature and therefore not really suitable for a single-imputation setting [13].

6 Conclusion

Congenial model combinations do not outperform uncongenial ones in settings with univariate missingness in the training data. For prediction models with continuous outcomes, neither performance nor calibration appears to be correlated with congeniality. While the interaction between the imputation model and the substantive prediction model can influence predictive performance, this effect is not driven by their alignment in model assumptions.

AI-assistance disclosure

This work benefited from the use of large-language models and AI coding tools. ChatGPT (OpenAI)[22] was used to support literature searching and to refine sentence grammar and style. Both ChatGPT and GitHub Copilot [23] were used to assist in coding and in the generation of figures and tables. All AI-generated content was carefully reviewed, edited, and approved to ensure accuracy and completeness.

References

[1] Manja Deforth, Georg Heinze, and Ulrike Held. The performance of prognostic models depended on the choice of missing value imputation algorithm: a simulation study. *Journal of Clinical Epidemiology*, 176:111539, 12 2024.

[2] Sarah Lee. Evaluating the impact of mean imputation on predictive models, March 2025. Accessed: 2025-04-28.

[3] C.L. Andaur Navarro, J.A.A. Damen, M. van Smeden, T. Takada, S.W.J. Nijman, P. Dhiman, J. Ma, G.S. Collins, R. Bajpai, R.D. Riley, K.G.M. Moons, and L. Hooft. Systematic review identifies the design and methodological conduct of studies on machine learning-based prediction models. *Journal of Clinical Epidemiology*, 154:8–22, Feb 2023. Epub 2022 Nov 25.

[4] Stef Van Buuren. *Flexible imputation of missing data*. Crc Press, Taylor & Francis Group, 2 edition, 2018.

[5] Shangzhi Hong and Henry S Lynn. Accuracy of random-forest-based imputation of missing data in the presence of non-normality, non-linearity, and interaction. *BMC Medical Research Methodology*, 20(1):199, 2020.

[6] JiaHang Li, ShuXia Guo, RuLin Ma, Jia He, XiangHui Zhang, DongSheng Rui, YuSong Ding, Yu Li, LeYao Jian, Jing Cheng, and Heng Guo. Comparison of the effects of imputation methods for missing data in predictive modelling of cohort study datasets. *BMC Medical Research Methodology*, 24, 02 2024.

[7] Xiao-Li Meng. Multiple-imputation inferences with uncongenial sources of input. *Statistical Science*, 9, 11 1994.

[8] Jonathan W Bartlett and Rachael A Hughes. Bootstrap inference for multiple imputation under uncongeniality and misspecification. *Statistical Methods in Medical Research*, 29:3533–3546, 06 2020.

[9] Carroll Orlagh. *Strategies for imputing missing covariate values in observational data.* PhD thesis, 01 2022.

[10] Tim P. Morris, Ian R. White, and Michael J. Crowther. Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, 38:2074–2102, 01 2019.

[11] Ewout W. Steyerberg. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Springer, New York, NY, 1 edition, 2009. See Chapter 13 for simulation and validation recommendations.

[12] Stef van Buuren and Karin Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 45(3):1–67, 2011.

[13] Rose Sisk, Matthew Sperrin, Niels Peek, Maarten van Smeden, and Glen Philip Martin. Imputation and missing indicators for handling missing data in the development and deployment of clinical prediction models: A simulation study. *Statistical Methods in Medical Research*, 32(8):1461–1477, 2023.

[14] Philipp Probst, Anne-Laure Boulesteix, and Bernd Bischl. Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(3):e1301, 2019.

[15] Marco Di Zio and Ugo Guarnera. Semiparametric predictive mean matching. *AStA Advances in Statistical Analysis*, 93:175–186, 10 2008.

[16] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.

[17] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2024.

[18] Alistair E. W. Johnson, Tom J. Pollard, Lu Shen, Li-Wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. Mimic-iii, a freely accessible critical care database. *Scientific Data*, 3:160035, 2016.

[19] Alistair E. W. Johnson, Tom J. Pollard, and Roger G. Mark. MIMIC-III Clinical Database (Version 1.4), 2016.

[20] Trent D. Buskirk and Stanislav Kolenikov. Finding respondents in the forest: A comparison of logistic regression and random forest models for response propensity weighting and stratification. *Survey Methods: Insights from the Field*, pages 1–17, 2015.

[21] Michikazu Nakai, Ding-Geng Chen, Kunihiro Nishimura, and Yoshihiro Miyamoto. Comparative study of four methods in missing value imputations

under missing completely at random mechanism. *Open Journal of Statistics*, 4(1):27–37, 2014.

[22] OpenAI. Chatgpt: Large language model. https://chat.openai.com/, 2023. Accessed: 2025-05-11.

[23] GitHub, Inc. GitHub Copilot. https://github.com/features/copilot, 2023. Accessed: 2025-05-11.

Appendix A

Appendix A contains a table that describes the different scenarios (Table 11, tables that show the mean RMSE (+standard deviation) of each model combination (Table 12 and Table 13) and tables that show the mean R^2 (+standard deviation) of each model combination (Table 14 and Table 15).

Table 11: This table helps to understand the following tables of Appendix A. It defines what data-generating mechanisms were used for each specific scenario. The "Scenario Index" is then used in the following tables (Table 12, Table 13, Table 14 and Table 15). For an accurate description of the data generating mechanisms see subsection 2.2.

Scenario Index	Relationship-Type Between Outcome	Relationship-Type Among Predictors	Relationship-Strength Among Predictors	Missing Mechanism
1	Quadratic	Quadratic	Low	MCAR
2	Quadratic	Quadratic	Low	Weak MAR
3	Quadratic	Quadratic	Low	Strong MAR
4	Quadratic	Quadratic	Low	Weak MNAR
5	Quadratic	Quadratic	Low	Strong MNAR
6	Quadratic	Quadratic	High	MCAR
7	Quadratic	Quadratic	High	Weak MAR
8	Quadratic	Quadratic	High	Strong MAR
9	Quadratic	Quadratic	High	Weak MNAR
10	Quadratic	Quadratic	High	Strong MNAR
11	Quadratic	Linear	Low	MCAR
12	Quadratic	Linear	Low	Weak MAR
13	Quadratic	Linear	Low	Strong MAR
14	Quadratic	Linear	Low	Weak MNAR
15	Quadratic	Linear	Low	Strong MNAR
16	Quadratic	Linear	High	MCAR
17	Quadratic	Linear	High	Weak MAR
18	Quadratic	Linear	High	Strong MAR
19	Quadratic	Linear	High	Weak MNAR
20	Quadratic	Linear	High	Strong MNAR
21	Linear	Quadratic	Low	MCAR
22	Linear	Quadratic	Low	Weak MAR
23	Linear	Quadratic	Low	Strong MAR
24	Linear	Quadratic	Low	Weak MNAR
25	Linear	Quadratic	Low	Strong MNAR
26	Linear	Quadratic	High	MCAR
27	Linear	Quadratic	High	Weak MAR
28	Linear	Quadratic	High	Strong MAR
29	Linear	Quadratic	High	Weak MNAR
30	Linear	Quadratic	High	Strong MNAR
31	Linear	Linear	Low	MCAR
32	Linear	Linear	Low	Weak MAR
33	Linear	Linear	Low	Strong MAR
34	Linear	Linear	Low	Weak MNAR
35	Linear	Linear	Low	Strong MNAR
36	Linear	Linear	High	MCAR
37	Linear	Linear	High	Weak MAR
38	Linear	Linear	High	Strong MAR
39	Linear	Linear	High	Weak MNAR
40	Linear	Linear	High	Strong MNAR

Table 12: The mean RMSE (+standard deviation) of each model combination with a regression prediction model across all 100 iterations. The scenario index (SI) represents different data generating mechanisms of each scenario (see Table 11). This table contains the following abbreviations: Predictive Mean Matching (PMM), Predictive Mean Matching with quadractic terms (PMM-Q), Regression Imputation (RI), Regression Imputation with quadratic terms (RI-Q), Random Forest (RF) and regression prediction model (REG).

SI	**PMM+REG**		**PMM-Q+REG**		**RI+REG**		**RI-Q+REG**		**RF+REG**	
	RMSE	**SD**	**RMSE**	**SD**	**RMSE**	**SD**	**RMSE**	**SD**	**RMSE**	**SD**
1	1.082	0.041	1.077	0.038	1.111	0.049	1.077	0.056	1.097	0.058
2	1.089	0.045	1.083	0.042	1.120	0.072	1.071	0.042	1.104	0.064
3	1.119	0.056	1.122	0.058	1.155	0.088	1.087	0.070	1.109	0.052
4	1.102	0.042	1.098	0.043	1.186	0.078	1.147	0.106	1.141	0.060
5	1.150	0.054	1.136	0.051	1.265	0.063	1.251	0.151	1.200	0.074
6	1.285	0.396	1.179	0.269	2.137	1.093	1.398	0.671	1.904	0.639
7	1.323	0.467	1.247	0.353	2.296	1.169	1.434	0.712	2.040	0.756
8	1.439	0.530	1.301	0.459	2.467	1.200	1.458	0.679	2.270	0.819
9	1.545	0.582	1.361	0.362	2.812	1.265	1.704	0.768	2.360	0.685
10	1.802	0.817	1.570	0.587	3.211	1.333	1.882	0.888	3.075	0.849
11	1.142	0.054	1.061	0.033	1.059	0.030	1.060	0.031	1.069	0.038
12	1.141	0.043	1.079	0.031	1.074	0.031	1.075	0.033	1.079	0.035
13	1.137	0.051	1.092	0.044	1.076	0.031	1.078	0.038	1.088	0.038
14	1.144	0.054	1.069	0.035	1.072	0.032	1.075	0.044	1.079	0.043
15	1.129	0.053	1.066	0.040	1.085	0.029	1.090	0.031	1.075	0.038
16	1.104	0.044	1.030	0.027	1.021	0.026	1.027	0.028	1.064	0.037
17	1.103	0.039	1.029	0.024	1.020	0.026	1.032	0.036	1.063	0.034
18	1.129	0.042	1.039	0.029	1.032	0.030	1.071	0.050	1.082	0.038
19	1.113	0.042	1.033	0.026	1.028	0.025	1.040	0.040	1.072	0.038
20	1.127	0.045	1.037	0.029	1.050	0.034	1.088	0.055	1.081	0.036
21	1.014	0.024	1.017	0.024	1.023	0.024	1.027	0.024	1.035	0.025
22	1.013	0.025	1.017	0.025	1.021	0.025	1.024	0.025	1.034	0.027
23	1.012	0.023	1.016	0.024	1.020	0.024	1.023	0.024	1.032	0.026
24	1.016	0.027	1.018	0.028	1.027	0.029	1.030	0.029	1.040	0.029
25	1.024	0.026	1.022	0.026	1.044	0.028	1.044	0.028	1.060	0.029
26	1.025	0.025	1.035	0.027	1.021	0.024	1.029	0.023	1.045	0.031
27	1.024	0.022	1.034	0.033	1.020	0.021	1.026	0.022	1.041	0.024
28	1.034	0.025	1.044	0.030	1.033	0.024	1.031	0.023	1.048	0.027
29	1.034	0.027	1.044	0.042	1.032	0.028	1.032	0.028	1.064	0.034
30	1.039	0.028	1.050	0.036	1.051	0.038	1.033	0.026	1.079	0.035
31	1.009	0.022	1.010	0.022	1.023	0.023	1.025	0.023	1.024	0.024
32	1.014	0.025	1.014	0.026	1.026	0.027	1.028	0.027	1.032	0.028
33	1.012	0.022	1.013	0.023	1.024	0.024	1.027	0.024	1.032	0.024
34	1.011	0.025	1.011	0.025	1.025	0.025	1.027	0.026	1.030	0.028
35	1.021	0.024	1.019	0.023	1.046	0.026	1.048	0.026	1.045	0.027
36	1.008	0.023	1.008	0.023	1.024	0.025	1.026	0.025	1.014	0.024
37	1.010	0.024	1.009	0.025	1.027	0.026	1.029	0.026	1.017	0.026
38	1.010	0.023	1.009	0.023	1.033	0.024	1.036	0.024	1.019	0.024
39	1.013	0.022	1.013	0.022	1.030	0.026	1.031	0.026	1.021	0.023
40	1.014	0.025	1.012	0.024	1.040	0.028	1.043	0.028	1.023	0.026

Table 13: The mean RMSE (+standard deviation) of each model combination with a random forest prediction model across all 100 iterations. The scenario index (SI) represents different data generating mechanisms of each scenario (see Table 11). This table contains the following abbreviations: Predictive Mean Matching (PMM), Predictive Mean Matching with quadractic terms (PMM-Q), Regression Imputation (RI), Regression Imputation with quadratic terms (RI-Q), Random Forest (RF) and random forest prediction model (RF).

SI	PMM+RF		PMM-Q+RF		RI+RF		RI-Q+RF		RF+RF	
	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD
1	1.642	0.138	1.641	0.138	1.672	0.133	1.583	0.131	1.650	0.150
2	1.654	0.126	1.646	0.130	1.675	0.121	1.587	0.125	1.657	0.147
3	1.694	0.150	1.693	0.140	1.731	0.132	1.622	0.142	1.689	0.167
4	1.668	0.134	1.663	0.125	1.722	0.117	1.608	0.127	1.684	0.146
5	1.742	0.127	1.747	0.140	1.837	0.126	1.693	0.126	1.785	0.164
6	3.195	1.854	3.384	1.911	3.253	1.828	3.217	1.880	3.659	1.926
7	3.292	1.832	3.435	1.948	3.296	1.797	3.300	1.861	3.762	1.865
8	3.233	1.742	3.336	1.874	3.265	1.697	3.190	1.761	3.718	1.787
9	3.596	1.964	3.793	2.031	3.643	1.911	3.666	1.963	4.107	1.940
10	3.588	1.923	3.839	2.062	3.711	1.810	3.547	2.021	4.038	1.884
11	1.581	0.104	1.486	0.102	1.506	0.098	1.505	0.096	1.516	0.098
12	1.563	0.099	1.474	0.086	1.499	0.092	1.499	0.093	1.492	0.092
13	1.574	0.094	1.483	0.103	1.506	0.097	1.505	0.099	1.501	0.100
14	1.587	0.097	1.499	0.092	1.522	0.090	1.521	0.089	1.518	0.095
15	1.589	0.100	1.494	0.102	1.533	0.103	1.527	0.103	1.515	0.103
16	1.519	0.177	1.450	0.185	1.431	0.181	1.438	0.186	1.479	0.179
17	1.520	0.196	1.464	0.210	1.439	0.201	1.453	0.202	1.485	0.194
18	1.515	0.132	1.462	0.147	1.426	0.131	1.466	0.133	1.470	0.136
19	1.519	0.149	1.454	0.166	1.434	0.150	1.451	0.155	1.474	0.156
20	1.525	0.160	1.470	0.164	1.447	0.161	1.483	0.163	1.477	0.156
21	1.131	0.031	1.155	0.042	1.124	0.030	1.125	0.030	1.172	0.034
22	1.129	0.031	1.156	0.040	1.122	0.029	1.125	0.030	1.170	0.034
23	1.126	0.027	1.159	0.037	1.123	0.026	1.126	0.027	1.172	0.029
24	1.125	0.029	1.150	0.035	1.118	0.029	1.119	0.029	1.171	0.029
25	1.138	0.029	1.157	0.039	1.133	0.030	1.131	0.029	1.186	0.035
26	1.110	0.033	1.136	0.046	1.096	0.032	1.105	0.032	1.135	0.034
27	1.117	0.039	1.154	0.051	1.103	0.040	1.114	0.037	1.141	0.040
28	1.118	0.030	1.145	0.039	1.106	0.029	1.117	0.030	1.140	0.030
29	1.122	0.044	1.148	0.053	1.109	0.047	1.115	0.044	1.149	0.046
30	1.117	0.047	1.135	0.049	1.110	0.046	1.110	0.043	1.146	0.045
31	1.122	0.030	1.145	0.037	1.124	0.030	1.124	0.030	1.158	0.033
32	1.124	0.031	1.143	0.037	1.124	0.031	1.125	0.031	1.158	0.033
33	1.117	0.028	1.140	0.036	1.122	0.029	1.122	0.029	1.154	0.031
34	1.121	0.029	1.139	0.037	1.123	0.030	1.123	0.030	1.155	0.031
35	1.124	0.027	1.138	0.030	1.133	0.027	1.132	0.028	1.169	0.028
36	1.073	0.024	1.099	0.036	1.079	0.027	1.080	0.027	1.083	0.025
37	1.076	0.027	1.095	0.033	1.082	0.029	1.083	0.029	1.087	0.029
38	1.076	0.026	1.088	0.032	1.086	0.025	1.085	0.025	1.088	0.027
39	1.079	0.026	1.102	0.036	1.086	0.028	1.085	0.029	1.091	0.027
40	1.078	0.027	1.092	0.031	1.087	0.028	1.087	0.027	1.088	0.028

Table 14: The mean R^2 (+standard deviation) of each model combination with a regression prediction model across all 100 iterations. The scenario index (SI) represents different data generating mechanisms of each scenario (see Table 11). This table contains the following abbreviations: Predictive Mean Matching (PMM), Predictive Mean Matching with quadractic terms (PMM-Q), Regression Imputation (RI), Regression Imputation with quadratic terms (RI-Q), Random Forest (RF) and regression prediction model (REG).

SI	**PMM+REG**		**PMM-Q+REG**		**RI+REG**		**RI-Q+REG**		**RF+REG**	
	R^2	**SD**	R^2	**SD**	R^2	**SD**	R^2	**SD**	R^2	**SD**
1	0.932	0.007	0.932	0.007	0.928	0.009	0.930	0.010	0.929	0.009
2	0.932	0.008	0.932	0.008	0.928	0.011	0.932	0.009	0.929	0.010
3	0.929	0.008	0.929	0.009	0.925	0.014	0.931	0.011	0.929	0.009
4	0.929	0.009	0.930	0.009	0.921	0.015	0.923	0.019	0.922	0.012
5	0.926	0.010	0.928	0.010	0.920	0.012	0.915	0.028	0.918	0.014
6	0.990	0.007	0.991	0.005	0.969	0.028	0.986	0.016	0.975	0.017
7	0.989	0.010	0.989	0.009	0.967	0.027	0.984	0.025	0.973	0.021
8	0.988	0.011	0.988	0.013	0.961	0.032	0.984	0.019	0.966	0.024
9	0.986	0.011	0.988	0.008	0.953	0.034	0.979	0.023	0.965	0.020
10	0.981	0.022	0.985	0.017	0.944	0.031	0.977	0.025	0.942	0.028
11	0.877	0.015	0.896	0.011	0.897	0.011	0.897	0.011	0.893	0.012
12	0.875	0.014	0.892	0.010	0.894	0.010	0.893	0.010	0.889	0.011
13	0.876	0.015	0.891	0.011	0.893	0.011	0.893	0.011	0.889	0.012
14	0.877	0.013	0.896	0.010	0.895	0.010	0.894	0.011	0.891	0.011
15	0.877	0.014	0.895	0.011	0.890	0.011	0.888	0.012	0.890	0.011
16	0.951	0.007	0.958	0.005	0.959	0.005	0.958	0.005	0.955	0.006
17	0.952	0.006	0.958	0.005	0.959	0.005	0.958	0.005	0.955	0.006
18	0.949	0.006	0.957	0.005	0.958	0.005	0.955	0.006	0.954	0.005
19	0.950	0.006	0.957	0.005	0.958	0.005	0.957	0.007	0.954	0.006
20	0.949	0.006	0.957	0.005	0.957	0.005	0.953	0.006	0.953	0.006
21	0.837	0.011	0.836	0.011	0.835	0.011	0.834	0.011	0.831	0.012
22	0.838	0.010	0.837	0.010	0.838	0.009	0.836	0.010	0.833	0.011
23	0.838	0.010	0.836	0.010	0.837	0.010	0.835	0.010	0.832	0.010
24	0.835	0.011	0.833	0.011	0.833	0.011	0.832	0.011	0.829	0.012
25	0.835	0.011	0.834	0.011	0.832	0.011	0.832	0.011	0.827	0.011
26	0.893	0.010	0.891	0.010	0.895	0.010	0.893	0.010	0.889	0.011
27	0.895	0.009	0.893	0.011	0.897	0.009	0.894	0.009	0.891	0.009
28	0.894	0.010	0.892	0.009	0.896	0.009	0.893	0.010	0.890	0.010
29	0.893	0.010	0.891	0.012	0.895	0.010	0.894	0.010	0.887	0.012
30	0.892	0.010	0.890	0.010	0.893	0.010	0.893	0.010	0.882	0.012
31	0.868	0.007	0.868	0.007	0.866	0.008	0.865	0.008	0.865	0.008
32	0.868	0.008	0.868	0.008	0.866	0.008	0.865	0.008	0.864	0.008
33	0.867	0.007	0.867	0.008	0.865	0.008	0.864	0.008	0.862	0.008
34	0.868	0.008	0.868	0.008	0.866	0.008	0.865	0.008	0.864	0.009
35	0.867	0.007	0.867	0.007	0.863	0.008	0.863	0.008	0.863	0.008
36	0.931	0.004	0.931	0.004	0.929	0.004	0.929	0.004	0.930	0.004
37	0.931	0.004	0.931	0.004	0.929	0.004	0.929	0.004	0.930	0.004
38	0.931	0.004	0.931	0.004	0.928	0.004	0.928	0.004	0.930	0.004
39	0.931	0.004	0.931	0.004	0.929	0.004	0.929	0.004	0.930	0.004
40	0.931	0.004	0.931	0.004	0.928	0.004	0.928	0.004	0.930	0.004

Table 15: The mean R^2 (+standard deviation) of each model combination with a random forest prediction model across all 100 iterations. The scenario index (SI) represents different data generating mechanisms of each scenario (see Table 11). This table contains the following abbreviations: Predictive Mean Matching (PMM), Predictive Mean Matching with quadractic terms (PMM-Q), Regression Imputation (RI), Regression Imputation with quadratic terms (RI-Q), Random Forest (RF) and random forest prediction model (RF).

SI	**PMM+RF**		**PMM-Q+RF**		**RI+RF**		**RI-Q+RF**		**RF+RF**	
	R^2	**SD**	R^2	**SD**	R^2	**SD**	R^2	**SD**	R^2	**SD**
1	0.843	0.019	0.843	0.018	0.834	0.019	0.851	0.018	0.846	0.020
2	0.843	0.016	0.845	0.018	0.836	0.018	0.852	0.016	0.846	0.019
3	0.837	0.020	0.837	0.019	0.825	0.020	0.846	0.021	0.842	0.022
4	0.836	0.018	0.837	0.016	0.826	0.017	0.847	0.016	0.837	0.019
5	0.827	0.017	0.826	0.021	0.815	0.020	0.838	0.018	0.823	0.024
6	0.937	0.057	0.931	0.059	0.935	0.057	0.939	0.055	0.925	0.062
7	0.935	0.058	0.931	0.061	0.936	0.053	0.939	0.055	0.924	0.057
8	0.934	0.059	0.931	0.065	0.934	0.057	0.938	0.060	0.922	0.060
9	0.922	0.066	0.915	0.070	0.921	0.064	0.924	0.065	0.909	0.064
10	0.924	0.062	0.914	0.071	0.923	0.058	0.932	0.062	0.913	0.057
11	0.770	0.025	0.794	0.020	0.790	0.023	0.790	0.023	0.788	0.022
12	0.772	0.020	0.794	0.017	0.789	0.018	0.789	0.018	0.791	0.018
13	0.770	0.019	0.794	0.015	0.788	0.016	0.788	0.016	0.789	0.016
14	0.771	0.022	0.793	0.018	0.788	0.018	0.788	0.018	0.789	0.019
15	0.764	0.027	0.790	0.020	0.778	0.023	0.780	0.022	0.785	0.021
16	0.916	0.013	0.923	0.014	0.924	0.013	0.923	0.014	0.920	0.013
17	0.916	0.015	0.922	0.017	0.924	0.016	0.922	0.016	0.920	0.015
18	0.915	0.011	0.922	0.012	0.925	0.010	0.921	0.011	0.921	0.011
19	0.915	0.012	0.922	0.014	0.923	0.012	0.922	0.012	0.920	0.012
20	0.913	0.015	0.921	0.014	0.921	0.014	0.918	0.015	0.919	0.013
21	0.799	0.015	0.791	0.018	0.800	0.014	0.800	0.014	0.789	0.016
22	0.802	0.013	0.792	0.016	0.802	0.012	0.801	0.012	0.792	0.014
23	0.801	0.012	0.790	0.015	0.800	0.012	0.799	0.012	0.790	0.013
24	0.800	0.012	0.790	0.015	0.800	0.012	0.800	0.012	0.787	0.012
25	0.798	0.012	0.791	0.017	0.798	0.013	0.799	0.013	0.786	0.014
26	0.875	0.011	0.869	0.013	0.878	0.010	0.876	0.010	0.870	0.011
27	0.875	0.010	0.867	0.012	0.878	0.010	0.876	0.010	0.871	0.010
28	0.875	0.010	0.869	0.012	0.877	0.010	0.875	0.010	0.871	0.010
29	0.874	0.011	0.869	0.013	0.877	0.011	0.876	0.011	0.869	0.012
30	0.875	0.013	0.871	0.013	0.876	0.013	0.876	0.012	0.869	0.013
31	0.840	0.009	0.833	0.011	0.837	0.010	0.837	0.010	0.833	0.010
32	0.841	0.010	0.835	0.011	0.839	0.010	0.838	0.010	0.834	0.010
33	0.839	0.010	0.833	0.012	0.836	0.010	0.836	0.009	0.832	0.011
34	0.840	0.010	0.835	0.012	0.838	0.010	0.838	0.010	0.834	0.011
35	0.840	0.010	0.835	0.011	0.836	0.010	0.836	0.010	0.832	0.010
36	0.922	0.004	0.918	0.006	0.921	0.005	0.921	0.005	0.921	0.005
37	0.922	0.005	0.919	0.005	0.921	0.005	0.921	0.005	0.921	0.005
38	0.922	0.004	0.920	0.005	0.920	0.004	0.921	0.004	0.921	0.005
39	0.922	0.005	0.919	0.006	0.921	0.005	0.921	0.005	0.921	0.005
40	0.922	0.005	0.920	0.005	0.921	0.005	0.921	0.005	0.921	0.005

Appendix B

Appendix B contains the calibration plots of all model combinations across all scenarios. The calibration plots are grouped into 8 distinct figures that based on their data-generating mechanism that align with the 8 plots of Figure 1. The plots are ordered as follows:

1. Relationship among predictors: Linear
   Correlation strength: Low
   Predictor-outcome relationship: Linear (see Figure 6).

2. Relationship among predictors: Linear
   Correlation strength: High
   Predictor-outcome relationship: Linear (see Figure 7).

3. Relationship among predictors: Quadratic
   Correlation strength: Low
   Predictor-outcome relationship: Linear (see Figure 8).

4. Relationship among predictors: Quadratic
   Correlation strength: High
   Predictor-outcome relationship: Linear (see Figure 9).

5. Relationship among predictors: Linear
   Correlation strength: Low
   Predictor-outcome relationship: Quadratic (see Figure 10).

6. Relationship among predictors: Linear
   Correlation strength: High
   Predictor-outcome relationship: Quadratic (see Figure 11).

7. Relationship among predictors: Quadratic
   Correlation strength: Low
   Predictor-outcome relationship: Quadratic (see Figure 12).

8. Relationship among predictors: Quadratic
   Correlation strength: High
   Predictor-outcome relationship: Quadratic (see Figure 13).

Figure 6: Calibration plots of model combinations in scenarios with linear relationship among predictors, low correlation strength among the predictors and linear relationship between predictors and outcome.

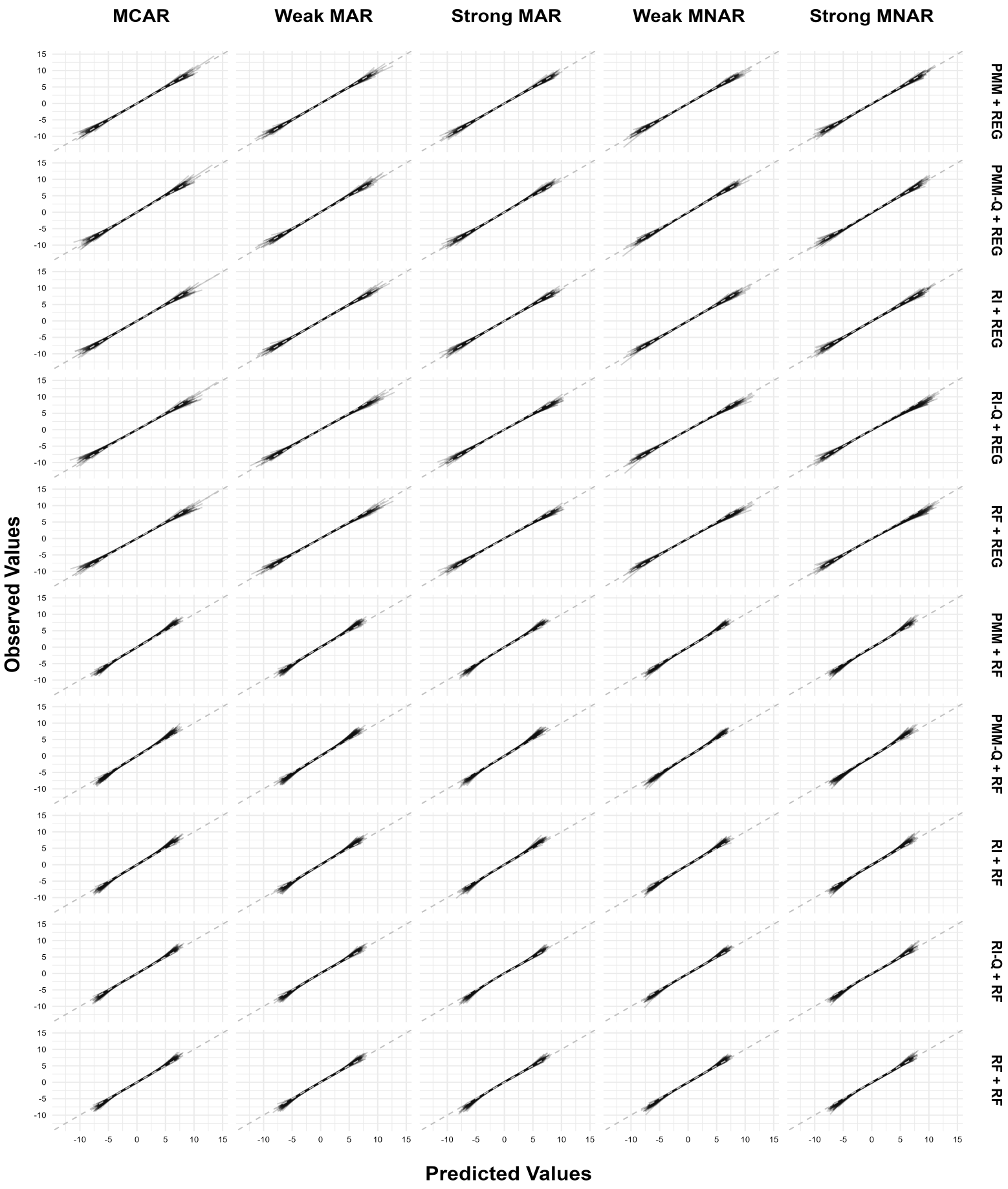


Figure 7: Calibration plots of model combinations in scenarios with linear relationship among predictors, high correlation strength among the predictors and linear relationship between predictors and outcome.

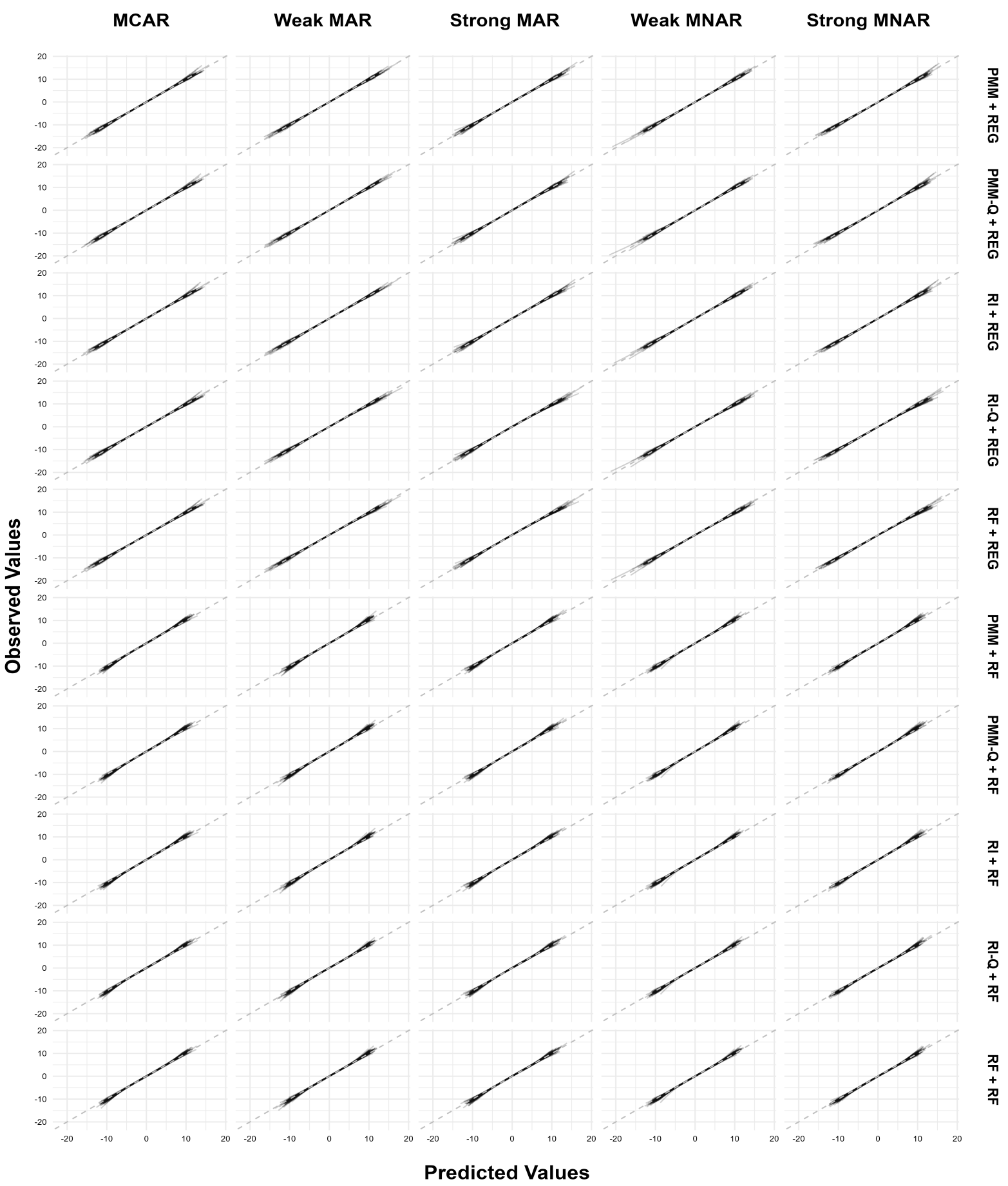


Figure 8: Calibration plots of model combinations in scenarios with quadratic relationship among predictors, low correlation strength among the predictors and linear relationship between predictors and outcome.

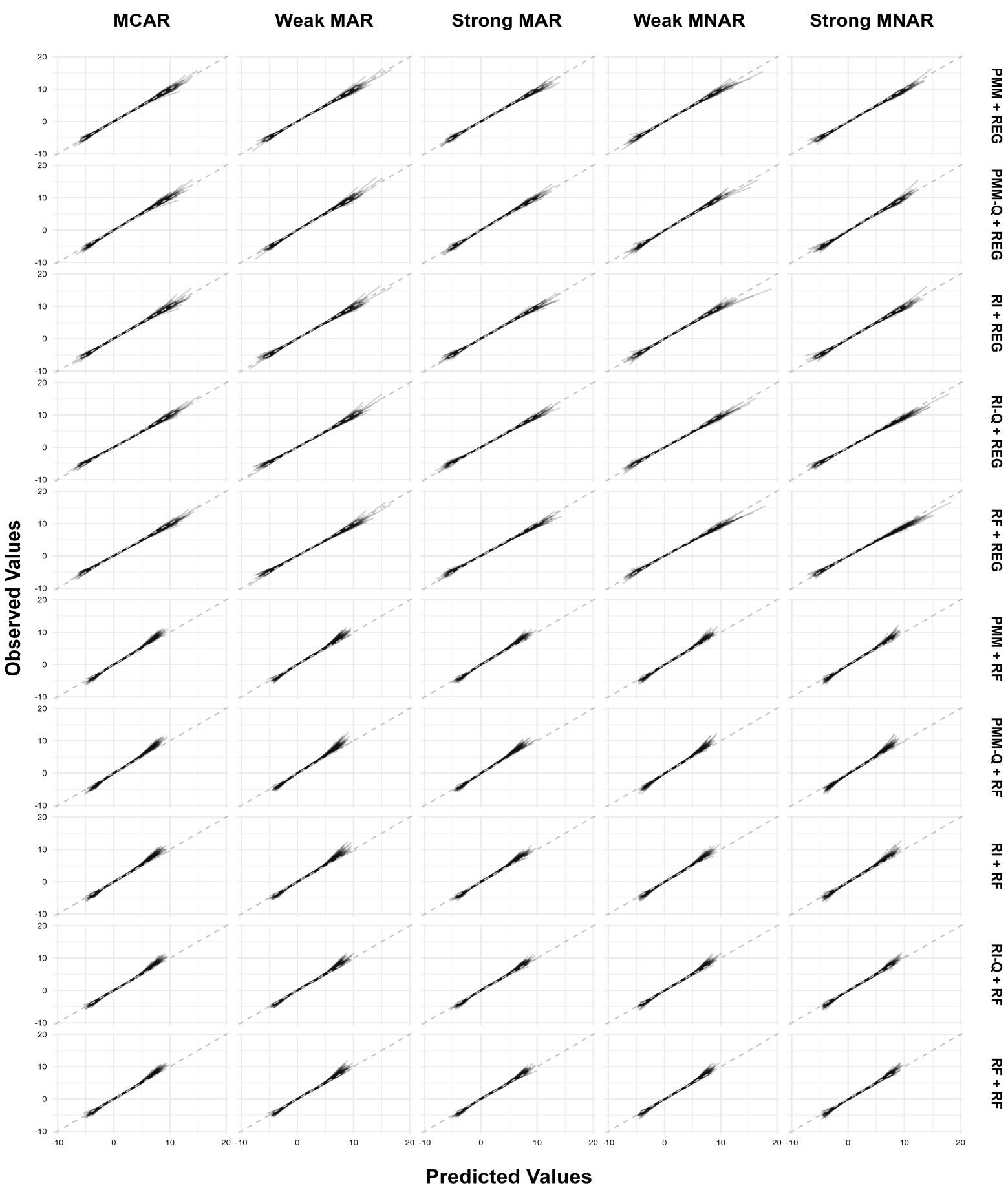


Figure 9: Calibration plots of model combinations in scenarios with quadratic relationship among predictors, high correlation strength among the predictors and linear relationship between predictors and outcome.

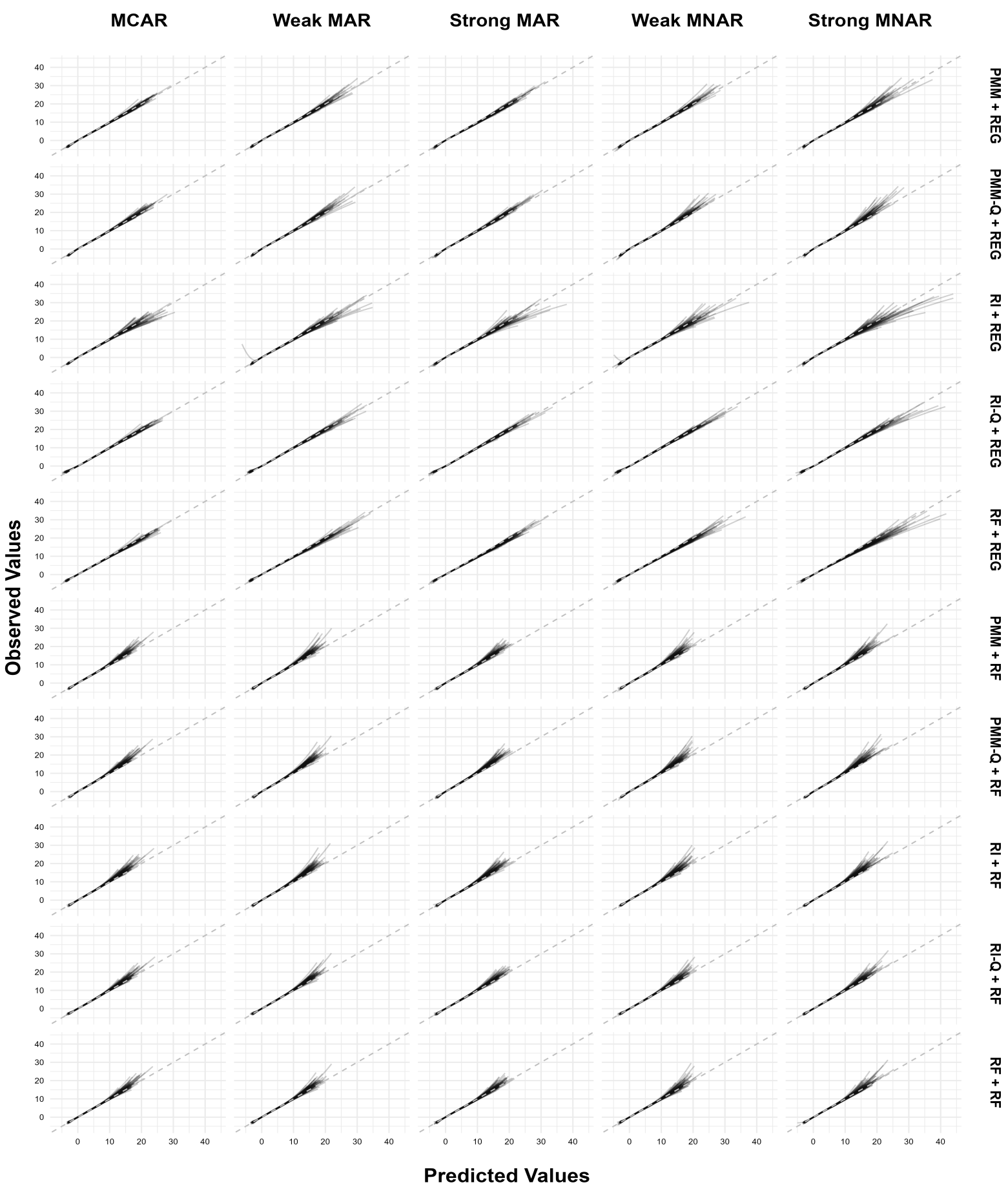


Figure 10: Calibration plots of model combinations in scenarios with linear relationship among predictors, low correlation strength among the predictors and quadratic relationship between predictors and outcome.

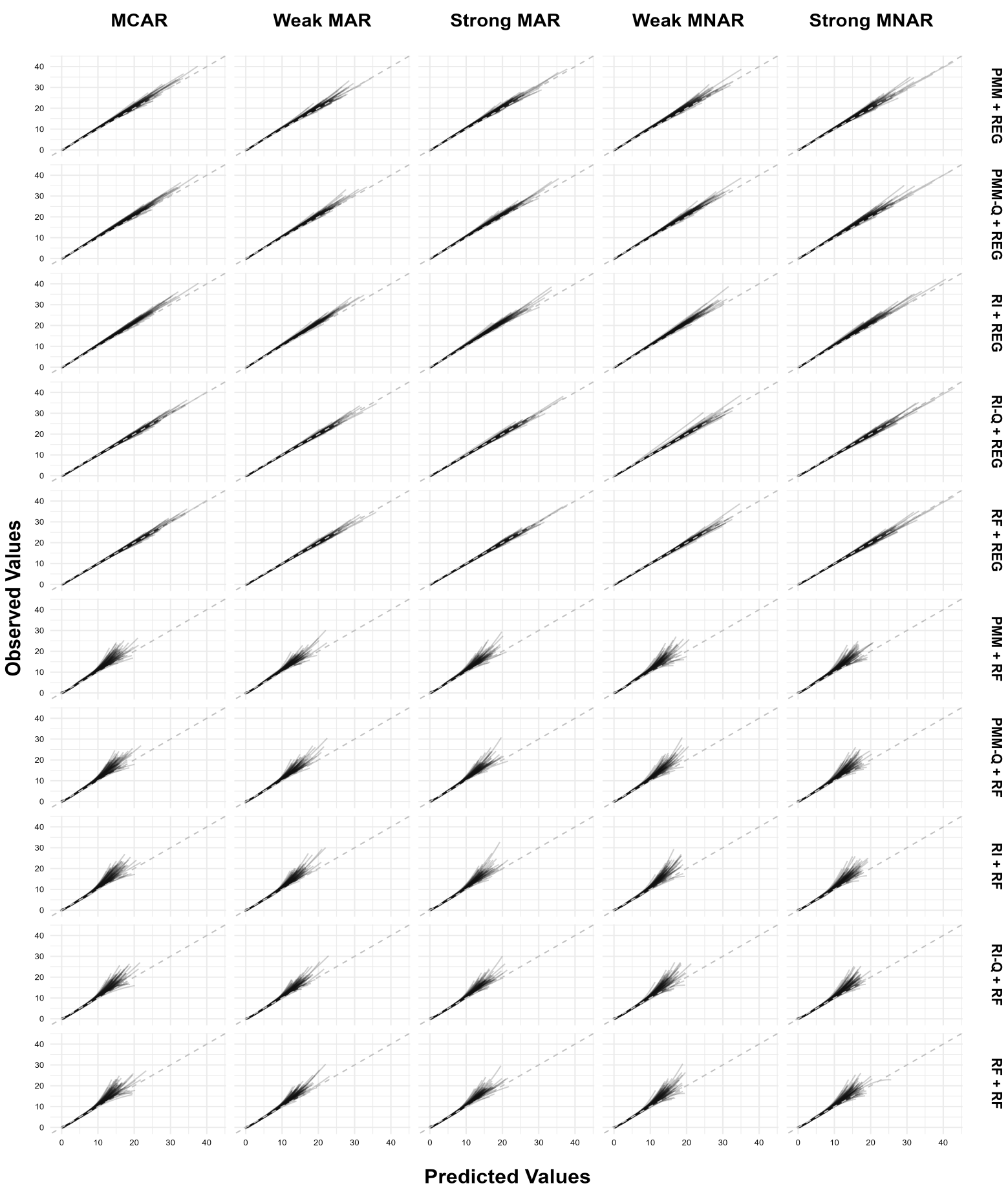


Figure 11: Calibration plots of model combinations in scenarios with linear relationship among predictors, high correlation strength among the predictors and quadratic relationship between predictors and outcome.

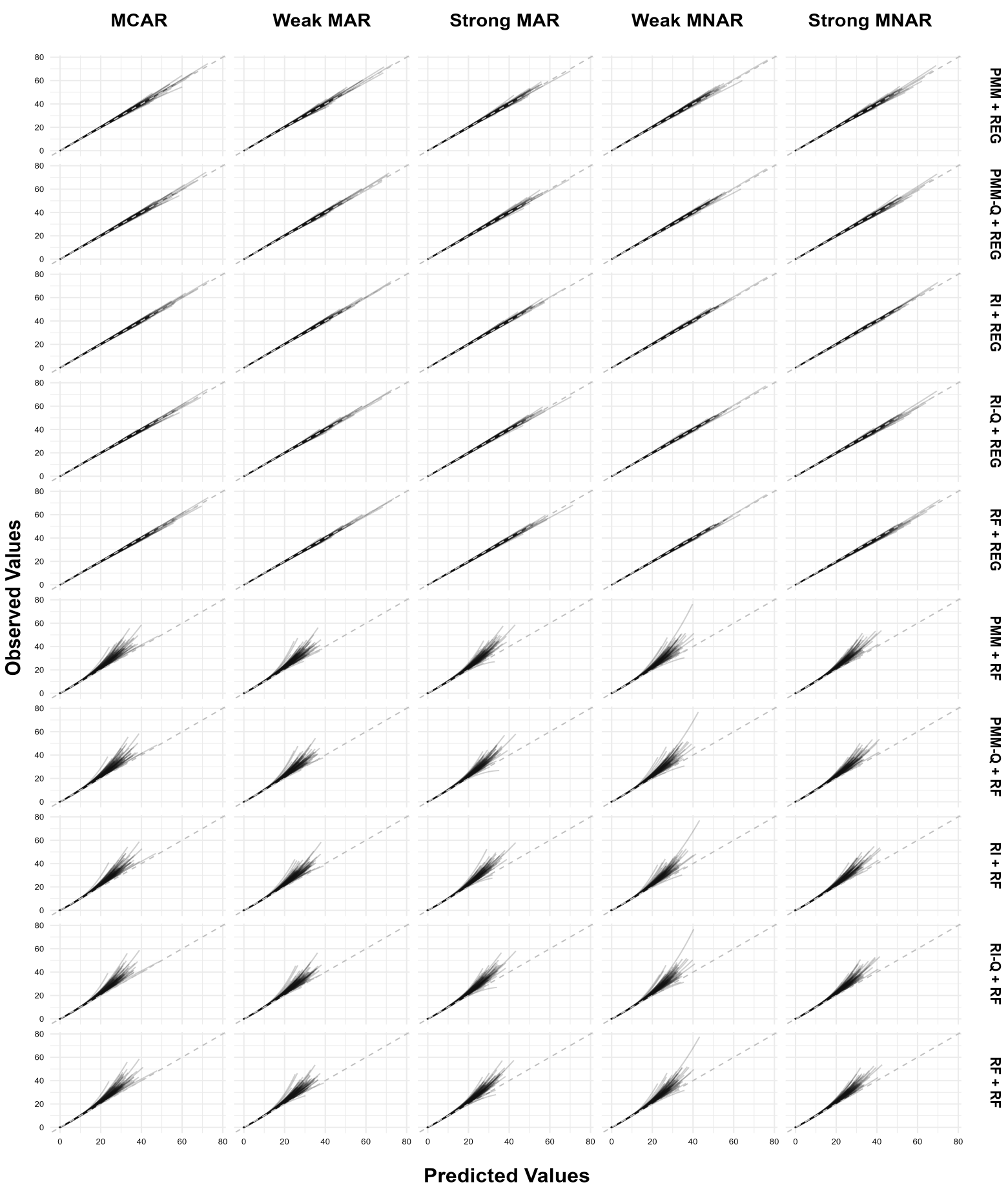


Figure 12: Calibration plots of model combinations in scenarios with quadratic relationship among predictors, low correlation strength among the predictors and quadratic relationship between predictors and outcome.

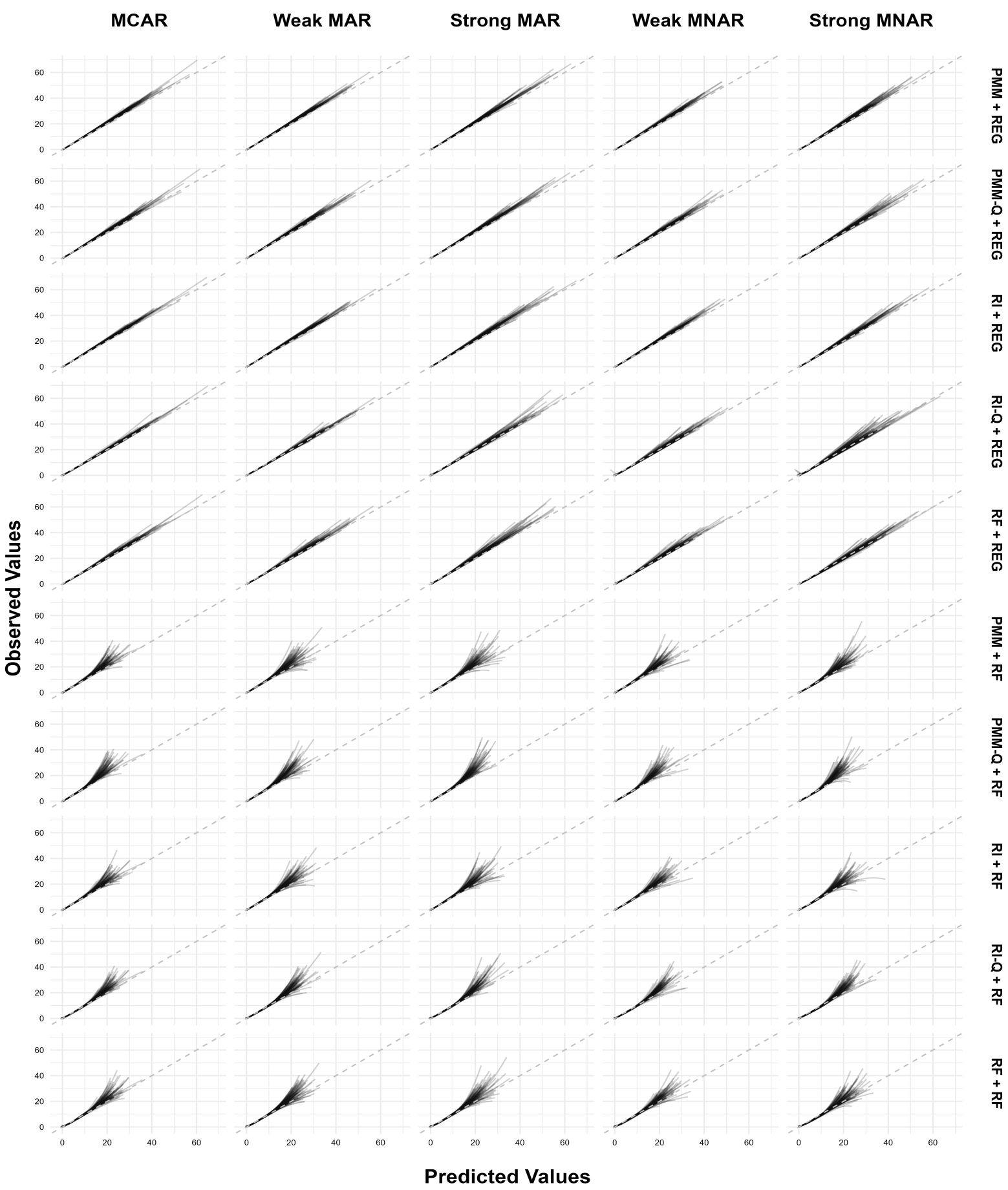
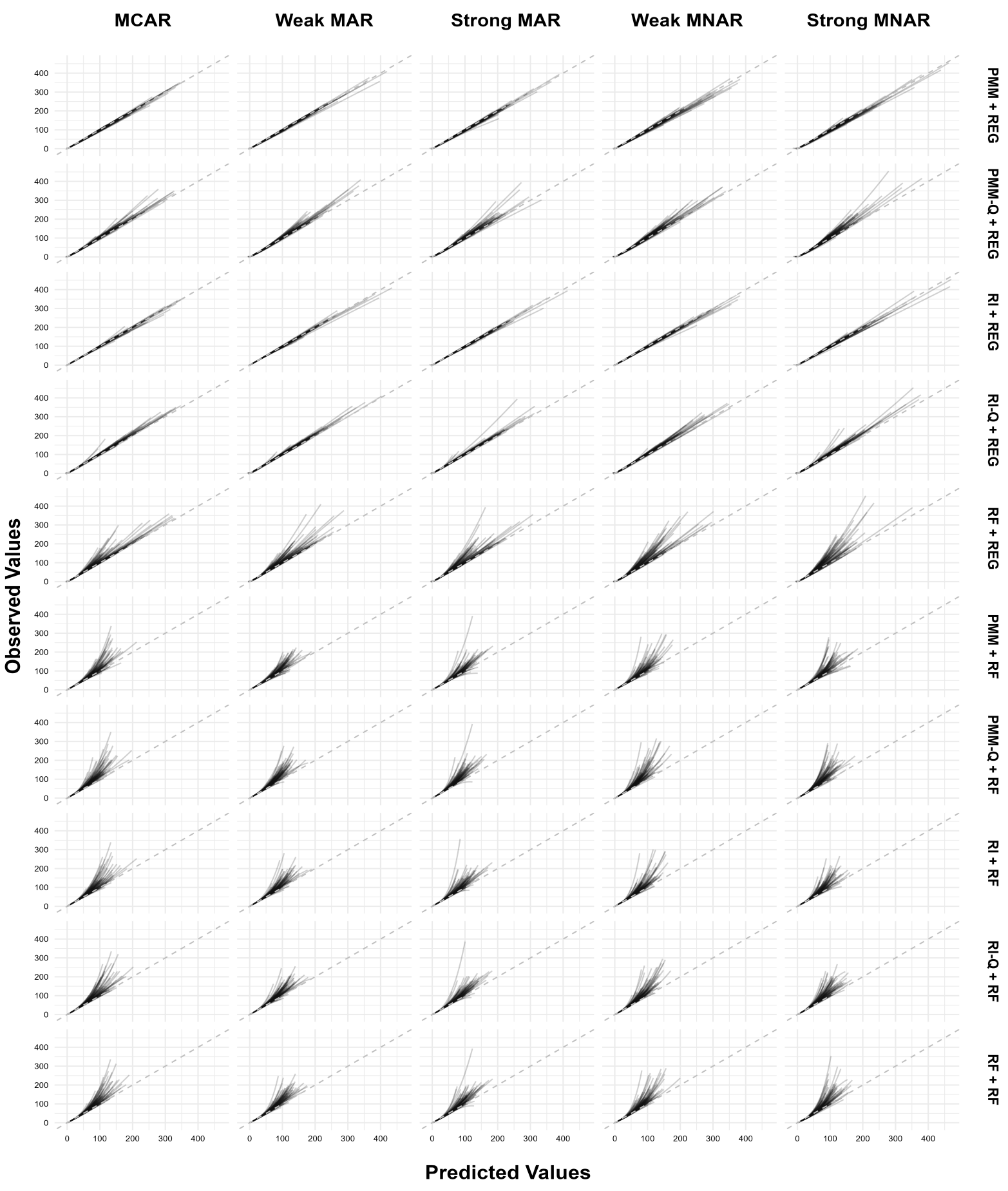


Figure 13: Calibration plots of model combinations in scenarios with quadratic relationship among predictors, high correlation strength among the predictors and quadratic relationship between predictors and outcome.



Appendix C

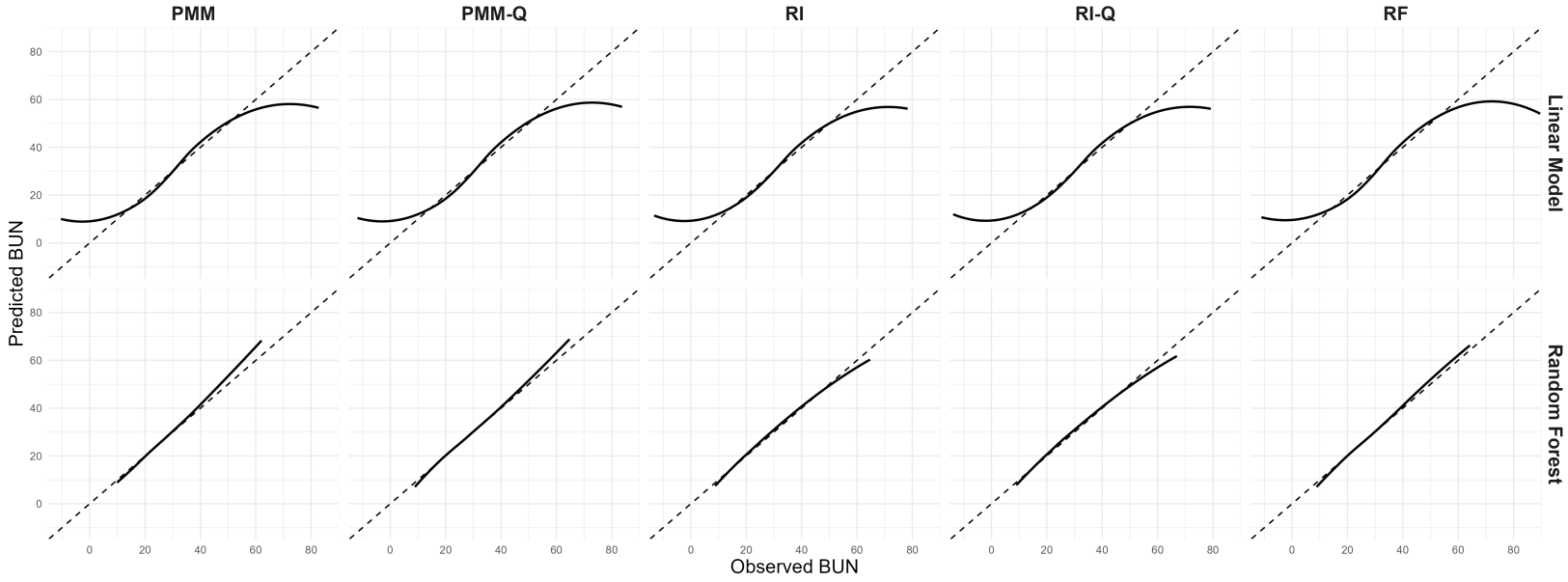


Figure 14: Calibration plots for the MIMIC applied example. This table contains the following abbreviations: Predictive Mean Matching (PMM), Predictive Mean Matching with quadractic terms (PMM-Q), Regression Imputation (RI), Regression Imputation with quadratic terms (RI-Q), Random Forest (RF).