Simulation Study Protocol

Merlin Urbanski

April 23, 2025

1 Introduction

In many scientific disciplines, especially in biomedical 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 essential measurements may be missing.

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). Although, it is advised against using these methods as they likely cause bias and loss of analytical power in prediction model development and in the predictive accuracy estimates [1] they still are still frequently used [2].

In recent years, imputation methods such as predictive mean matching (PMM), regression imputation (RI), and random forest (RF) imputation have become increasingly popular in prediction modeling. While there are already some simulation studies comparing the predictive performance of different imputation models they most often compare a variety of imputation models paired with one single substantive prediction model [3] [4].

However, focusing on the performance of individual imputation methods in isolation overlooks an important aspect: the interaction between the imputation model and the substantive prediction model. Recent research suggests that a phenomenon called uncongeniality—the opposite of congeniality and closely related to compatibility and incompatibility—can have a strong influence on model performance, particularly in the context of parameter estimation. Uncongeniality was first described by Meng [5] 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 (MCmbs) can lead to biased parameter estimates and invalid inferences [6] [7].

However, there is relatively little research on how uncongeniality affects performance in a predictive setting. As Orloagh ([8] 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 MCmbs. These included imputation methods from the mice package, combined with prediction models based on linear regression and random forests.

2 Methods

2.1 ADEMP

The study adheres to the ADEMP guidelines for the design and reporting of the simulation study [?].

2.2 Aim

We aim to determine the effect of congeniality between an imputation model and a substantive prediction model on predictive performance. Under 40 realistic scenarios of univariate missingness ten model combinations will be compared based on their out-of-sample predictive performance.

2.3 Data-Generating Mechanisms

2.3.1 Scenarios

Data with only continous predictors and a continous outcome will be simulated to reflect 40 (5 x 2 x 2 x 2) unique scenarios. This is achieved by varying the following four characteristics of the data: missingness-mechanism (MCmbAR, weak MAR, strong MAR, weak MNAR, strong MNAR), type of correlation between the one predictor with missingness and the predictors without missingness linear and quadratic, type of correlation between the predictors and the outcome variable linear and quadratic and strength of correlation between the predictors low (0.2) and high (0.8).

Table 1: Summary of factors to be varied in the data simulation. Data will be simulated to reflect 40 unique scenarios (5 x 2 x 2 x 2), by varying the following characteristics.

Factor	**Levels**
Missingness Mechanism	MCmbAR 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 (0.2), High (0.8)

All training data sets will have a sample size of 1000 observations while the test sets are 100 times larger and have a sample size of 100,000 observations. Univariate missingness in one of the predictors will occur only in the training data set and will be constant at 30% across all scenarios.

2.3.2 Data-Generating Mechanisms

For all scenarios data of 7 of the 8 predictors will be generated from a multivariate normal distribution.

For all scenarios, the data for the 7 predictors $X_2, \ldots, X_8$ are generated as follows:

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

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 $\rho \in \{0.2, 0.8\}$ depending on the simulation scenario.

2.3.3 Generation of X_1 with a Linear Relationship between Predictors

For the linear relationship scenarios, the predictor X_1 is constructed as a weighted sum of the other seven predictors X_2 to X_8 plus an additive noise term. The weights and noise are calibrated to ensure that, in the linear case, X_1 has the same correlation strength (either 0.2 or 0.8) with $X_2, \ldots, X_8$ as the correlations among these predictors.

1. **Linear Combination:**
   X_1 is defined as a scaled sum of the seven predictors:

   $$X_1 = a \cdot \sum_{j=2}^{8} X_j + \varepsilon,$$

   where a is a scaling factor and $\varepsilon \sim \mathcal{N}(0, \sigma_e^2)$ is the error term.

2. **Low Covariance Case ($cov = 0.2$):**
   In the low covariance scenario, the scaling factor is set to

   $$a = \frac{1}{11},$$

   and the error standard deviation is

   $$\sigma_e = \sqrt{0.6706},$$

   ensuring that the introduction of noise into X_1 results in X_1 correlating with $X_2, \ldots, X_8$ at approximately 0.2, thereby matching the inter-predictor correlations.

3. **High Covariance Case ($cov = 0.8$):**
   In the high covariance scenario, the scaling factor is adjusted to
   $$a = \frac{0.8}{5.8},$$
   and the error standard deviation becomes
   $$\sigma_e = \sqrt{0.228},$$
   ensuring that X_1 exhibits a correlation of approximately 0.8 with $X_2, \ldots, X_8$, in line with the stronger inter-predictor correlations.

By carefully choosing the scaling factor a and the noise level σ_e, the simulation ensures that the correlation between X_1 and the predictors X_2 to X_8 matches the inter-predictor correlations. As a result, in the high correlation scenario, $\mathrm{Var}(X_1)$ is approximately 1, while in the low correlation scenario it is around 0.8.

2.3.4 Generation of X_1 with a Quadratic Relationship between Predictors

But why are we generating X_1 in such a complicated way when we could have sampled all eight predictors from a multivariate normal distribution and avoiding the issue of differing variances for X_1? The reason is that we want to create a comparable scenario in which the association between the predictors $X_2, \ldots, X_8$ and X_1 is of similar strength as in the linear case, but the relationship is quadratic rather than linear. However, the options for directly comparing a linear to a non-linear relationship are limited. Therefore, we decided to simulate X_1 as a quadratic function of $X_2, \ldots, X_8$ with added noise.

1. **Quadratic Combination:**
   In order to create a quadratic relationship analogous to the linear case, we simply replace each predictor X_j with its square. Thus, X_1 is defined as:
   $$X_1 = a \cdot \sum_{j=2}^{8} X_j^2 + \varepsilon,$$
   where a is a scaling factor and $\varepsilon \sim \mathcal{N}(0, \sigma_e^2)$ is the error term.

2. **Low Covariance Case ($cov = 0.2$):**
For the low covariance scenario, we adopt the same structure as in the linear case, but now applied to the squared predictors. We take the exact same numbers from the linear sceanrio and therefore set

$$a = \frac{1}{11} \quad \text{and} \quad \sigma_e = \sqrt{0.6706},$$

ensuring that the resulting quadratic effect mirrors the low inter-predictor correlation.

3. **High Covariance Case ($cov = 0.8$):**
Similarly, for the high covariance scenario, we take the previously calculated values from the linear scenario and set

$$a = \frac{0.8}{5.8} \quad \text{and} \quad \sigma_e = \sqrt{0.228},$$

.

In essence, we extend the linear model by squaring the predictors to induce a quadratic effect while retaining the overall structure. This approach allows us to compare model performance under both linear and quadratic relationships on a similar scale. However, metrics that evaluate the strength of these relationships—such as R^2 in regression models and the mutual information criterion (MIC)—yield different results when comparing the relationships in this study. Specifically, R^2 suggests that the quadratic (squared) relationship is stronger than the linear one, whereas the MIC indicates that the linear relationship is stronger. Therefore, we believe that our approach allows for a fair comparison. Similarly to the linear scenario, the variance of X_1 is higher in the high-correlation scenario ($\text{Var}(X_1) \approx 1.5$) compared to the low-correlation scenario ($\text{Var}(X_1) \approx 0.8$).

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. Out of the 8 predictors, the first 2 are assumed to have a strong effect, the next 2 a weak effect, and the remaining 4 no direct effect on the outcome. Moreover, the effect on Y can be either linear or quadratic.

In our simulation, the outcome is generated as follows:

1. **Linear Effects:**
   When the relationship is linear, the outcome is constructed as a linear combination of the first four predictors with weights of 1.5 for the strong-effect predictors and 0.5 for the weak-effect predictors, plus normally distributed noise:
   $$Y = 1.5\,X_1 + 1.5\,X_2 + 0.5\,X_3 + 0.5\,X_4 + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0,1).$$

2. **Quadratic Effects:**
   When a quadratic relationship is assumed, we extend the linear model by replacing the predictors with their squares. Thus, the outcome is given by:
   $$Y = 1.5\,X_1^2 + 1.5\,X_2^2 + 0.5\,X_3^2 + 0.5\,X_4^2 + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0,1).$$

This design allows for a direct comparison of model performance when the true relationship between the predictors and the outcome is linear versus quadratic. The use of different weights for the strong and weak predictors, along with the additive noise, ensures that the generated outcome reflects a realistic and variable response.The generated scenarios are visualized in a simplified manner in Figure 1.

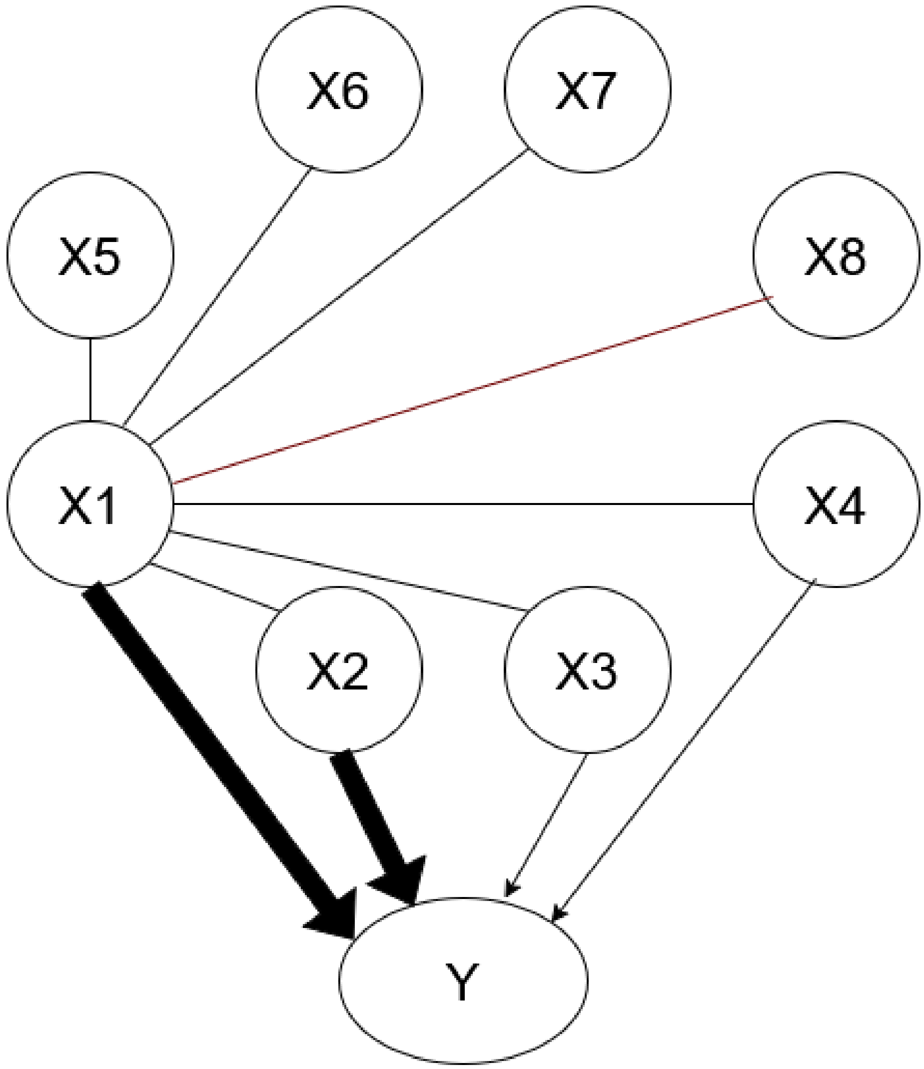

Figure 1: Visualization of the generated data. The lines among X1 and the other predictors are the relationships that can be linear or quadratic. The thick-dark arrows are the strong effects on the outcome variable Y and the thin-light arrows are the weak effects on Y.

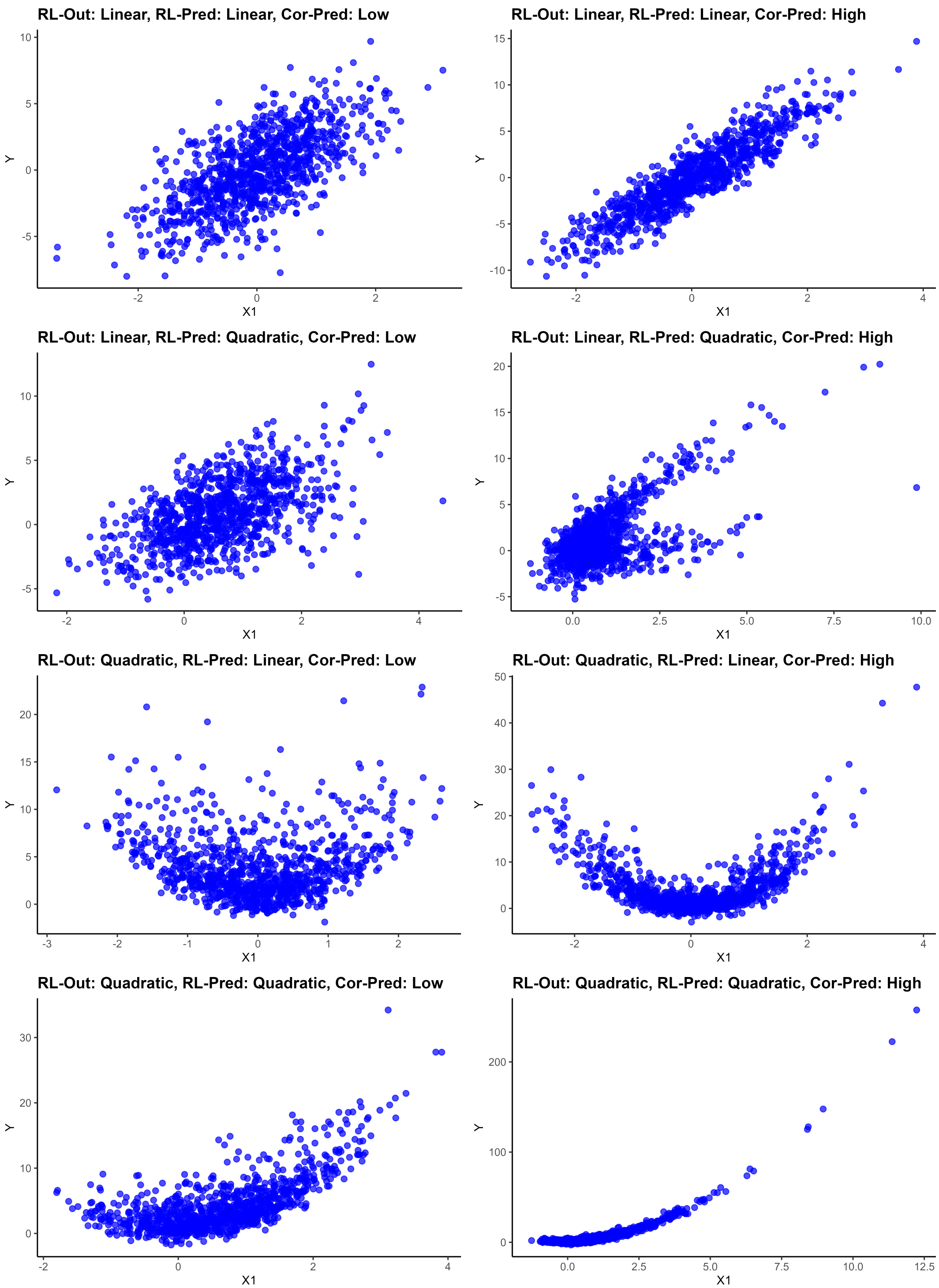

Figure 2: Visualization of the relationship between X1 and the outcome Y in 8 distinct scenarios. In the scenarios things varied: the type relationship between the predictors and outcome (RL-OUT), the type of relationship among predictors (RL-Pred) and strength of the relationship among predictors (Cor-Pred).

2.5 Missing Data Mechanisms

Missing data 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 in question.

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

   Under the MCmbAR 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. Mathematically, this is expressed as:

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

   where:

   - w is a weight (for example, 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. In this case, the probability that an observation is missing is given by:

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

   after which it is scaled appropriately so that the average probability matches the target missing percentage.

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

   In the weak MNAR scenario, the mechanism is similar to weak MAR but the dependency is on the variable X_1 itself. The missingness probability is computed as:

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

   where the same definitions apply: w moderates the contribution of the rank of X_1 relative to the random component.

5. **Strong MNAR:**

   Under the strong MNAR mechanism, the missingness is determined exclusively by X_1. The probability is defined by:

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

   .

In each case, once the probability is computed, it is adjusted so that its average equals the intended missing percentage. Finally, for every observation, a Bernoulli draw with the computed probability is performed to decide whether the value is missing ($M = 1$) or observed ($M = 0$).

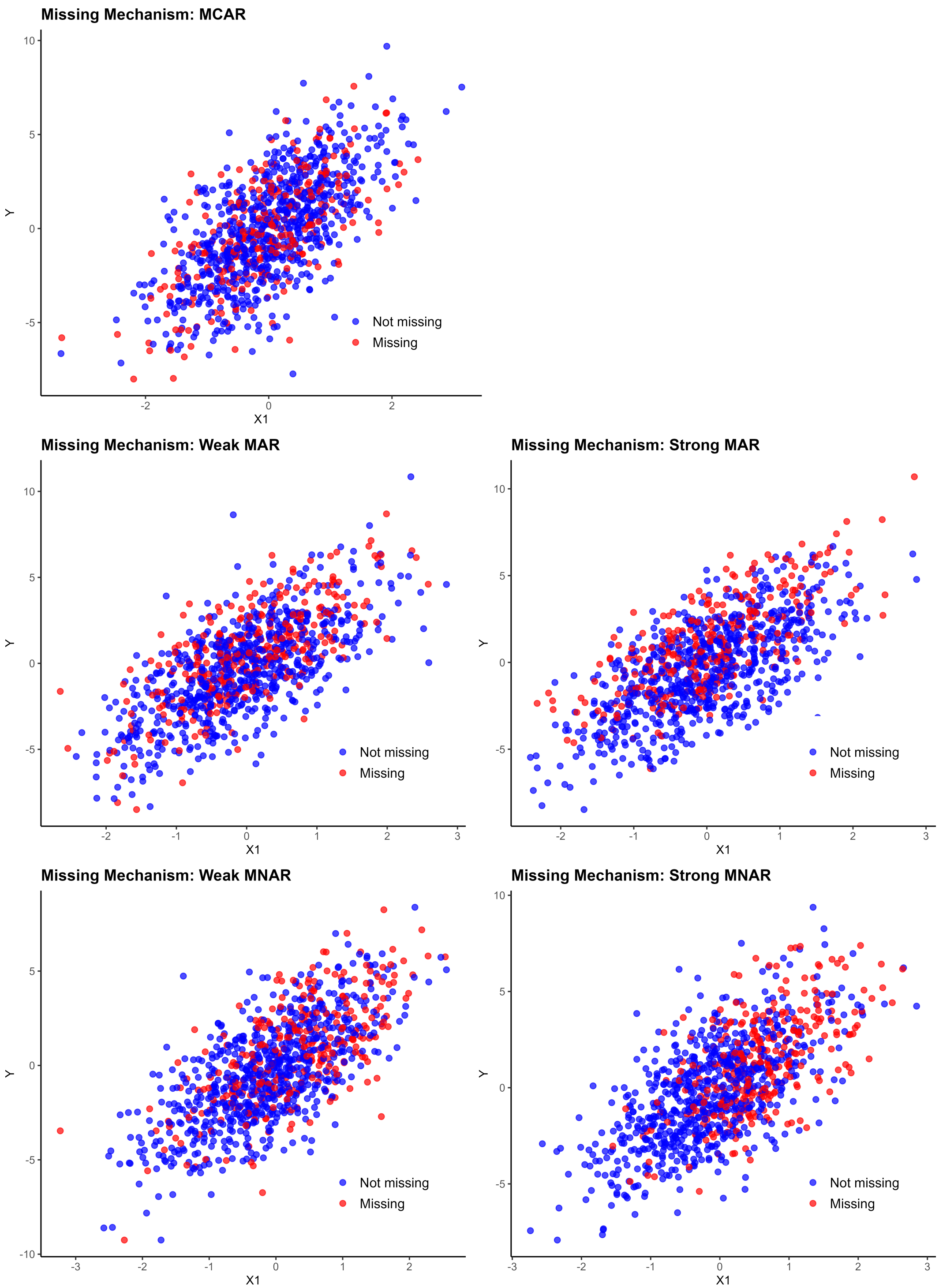

Figure 3: Visualization of the 5 Missing Mechanisms.

3 Estimands

The estimands in this study are metrics of predictive performance that evaluate the relationship between the predicted values ($\hat{Y}$) and the true outcome values (Y). These include:

1. **Root Mean Squared Error (RMSE):** RMSE evaluates the average size of error between the predicted and true values, providing a measure of overall prediction 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. **Visual Evaluation of 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 degree 1:1 line indicate systematic over- or underestimation, providing insights into the alignment of predictions with true outcomes.

These estimands were selected to assess the performance of congenial and uncongenial model combinations, evaluating different dimensions of predictive accuracy, explanatory power, and calibration.

4 Methods

4.1 Imputation Methods

To address the univariate missing data, we employed single imputation using chained equations via the R package `mice` [?]. Given our focus on predictive accuracy rather than inference, we opted against multiple imputations. For all methods, the number of iterations per imputation was set to the default value of 5.

We evaluated five imputation strategies to generate single imputed datasets for model training: (1) predictive mean matching (PMM) with one donor (mice::pmm), (2) PMM with one donor 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). For the PMM approaches, the number of donors was set to one. Additionally, we used regression imputation instead of the commonly used Bayesian regression imputation, as our primary objective was to achieve the most accurate imputed values for prediction, rather than to capture imputation uncertainty for inferential purposes.

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`

4.2 Prediction Models

In our study, we employed two distinct predictive modeling approaches, both of which were trained on the imputed training dataset and subsequently applied to the test set. The first approach was a regression model that incorporated both linear and quadratic terms for all predictors, enabling us to capture potential non-linear effects while maintaining model interpretability.

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 model evaluation. To fine-tune the random forest, 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 test dataset.

Table 3: Prediction Models Employed

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

4.3 Model Combinations

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 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 and a regression prediction model. 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 [**?**]. This results in PMM having less rigid assumptions than those of fully parametric regression, so the differences in assumptions are less pronounced.

Finally, there are two congenial model combinations in which the imputation model is essentially identical to the prediction model, sharing the same underlying assumptions.

Table 4: Model Combination Congeniality (using abbreviations)

Imputation Model	**Prediction Model**	**Congeniality**
PMM	Reg	Rather Uncongenial (RU)
PMM-Q	Reg	Rather Congenial (RC)
RI	Reg	Uncongenial (U)
RI-Q	Reg	Congenial (C)
RF	Reg	Uncongenial (U)
PMM	RF	Uncongenial (U)
PMM-Q	RF	Uncongenial (U)
RI	RF	Uncongenial (U)
RI-Q	RF	Uncongenial (U)
RF	RF	Congenial (C)

5 Results

5.1 RMSE and R^2 within different scenarios

Model performance varied substantially across the simulation scenarios. However, in nearly every case the missing data mechanism had little influence on RMSE (see Appendix A) or R^2(see Appendix B). Accordingly, we concentrated our analysis 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.

It is important to note that the variance of the outcome variable Y can differ strongly due to the different data generating mechanisms. Therefore, it is not possible to directly compare the RMSE between scenarios. Instead we focused on comparing the ranking of the different Mcmbs between scenarios.

5.1.1 Linear Relationships among Predictors and between Predictors and Outcome

When considering model 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. The MCmb including RF performed better in the high correlation scenario but overall the differences between model combinations are relatively small.

When examining combinations using a RF prediction model, those using PMM as the imputation method performed best, followed by combinations based on RI (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.

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.

		Low Correlation				High Correlation				
Imputation Model	**Prediction Model**	RMSE	SD	R^2	SD	RMSE	SD	R^2	SD	**Congeniality**
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

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

In low-correlation scenarios with regression prediction models, the results mirrored those observed previously with purely linear relationships. Both PMM-based methods performed best, followed closely by RI-based methods. Interestingly, methods including quadratic effects (PMM-Q, RI-Q) performed slightly worse than their linear-only counterparts (PMM, RI), despite the data containing quadratic relationships.

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.

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

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.

		Low Correlation				High Correlation				
Imputation Model	**Prediction Model**	RMSE	SD	R^2	SD	RMSE	SD	R^2	SD	**Congeniality**
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

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

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

For RF prediction models, the ranking of MCmbs 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, MCmbs using PMM consistently perform the worst.

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.

		Low Correlation				High Correlation				
Imputation Model	**Prediction Model**	RMSE	SD	R^2	SD	RMSE	SD	R^2	SD	**Congeniality**
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

5.1.4 Quadratic Relationships among Predictors and between Predictors and Outcome

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

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

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.

		Low Correlation				High Correlation				
Imputation Model	**Prediction Model**	RMSE	SD	R^2	SD	RMSE	SD	R^2	SD	**Congeniality**
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

5.2 Performance across all scenarios

Analysis of the RMSE and R^2 revealed that MCmbs 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 MCmb using a regression prediction model performed better than a MCmb using a RF prediction model.

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

The performance of MCmbs varied significantly across different scenarios. No single MCmb consistently ranked within the top two across all scenarios. Additionally, the effectiveness of imputation methods depended on the prediction model that it is combined with (see Table). However, the congeniality or uncongeniality of the MCmb did not predict whether a MCmb would perform well or poorly.

Table 9: Mean rank of RMSE and mean rank of R^2 by MCmb 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.

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

5.3 Calibration Plots

Similar to other performance metrics, missing data mechanisms had minimal impact on the calibration plots (see Appendix C). We therefore focus on comparing the MCAR scenarios.

Across all scenarios, calibration curves of MCmbs 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.

In scenarios involving quadratic effects either among predictors or between predictors and outcome, calibration curves of MCmbs with RF prediction models showed a notable spread at higher predicted values. This spread became more pronounced under conditions of high predictor correlation. Furthermore, these calibration curves consistently lay above the 45-degree 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-degree line, indicating systematic overestimation.

All other MCmbs showed relatively good calibration with only minor deviations from the 45-degree line.

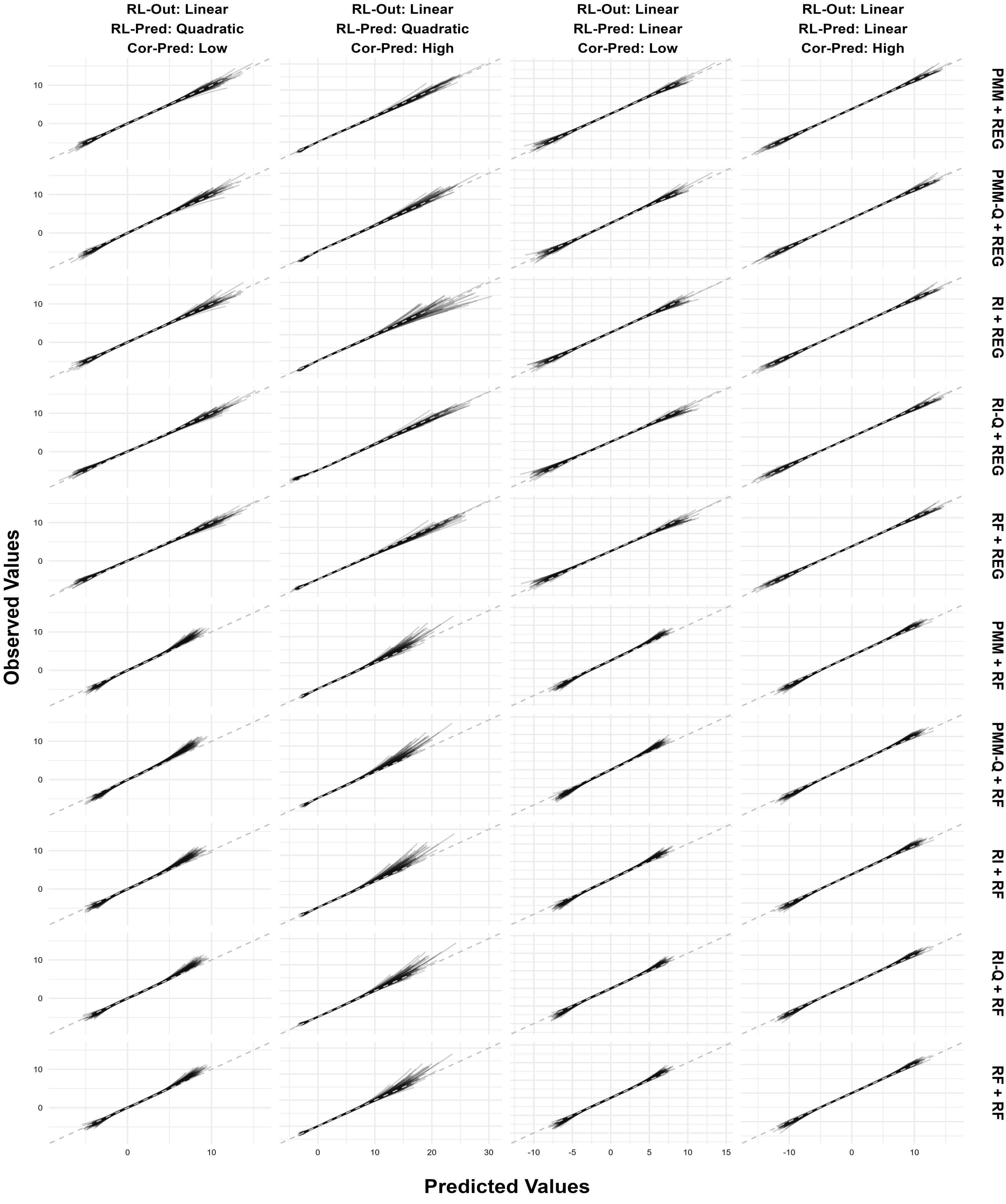

Figure 4: Calibration Plots of different scenarios with a linear outcome-relationship (all MCAR).

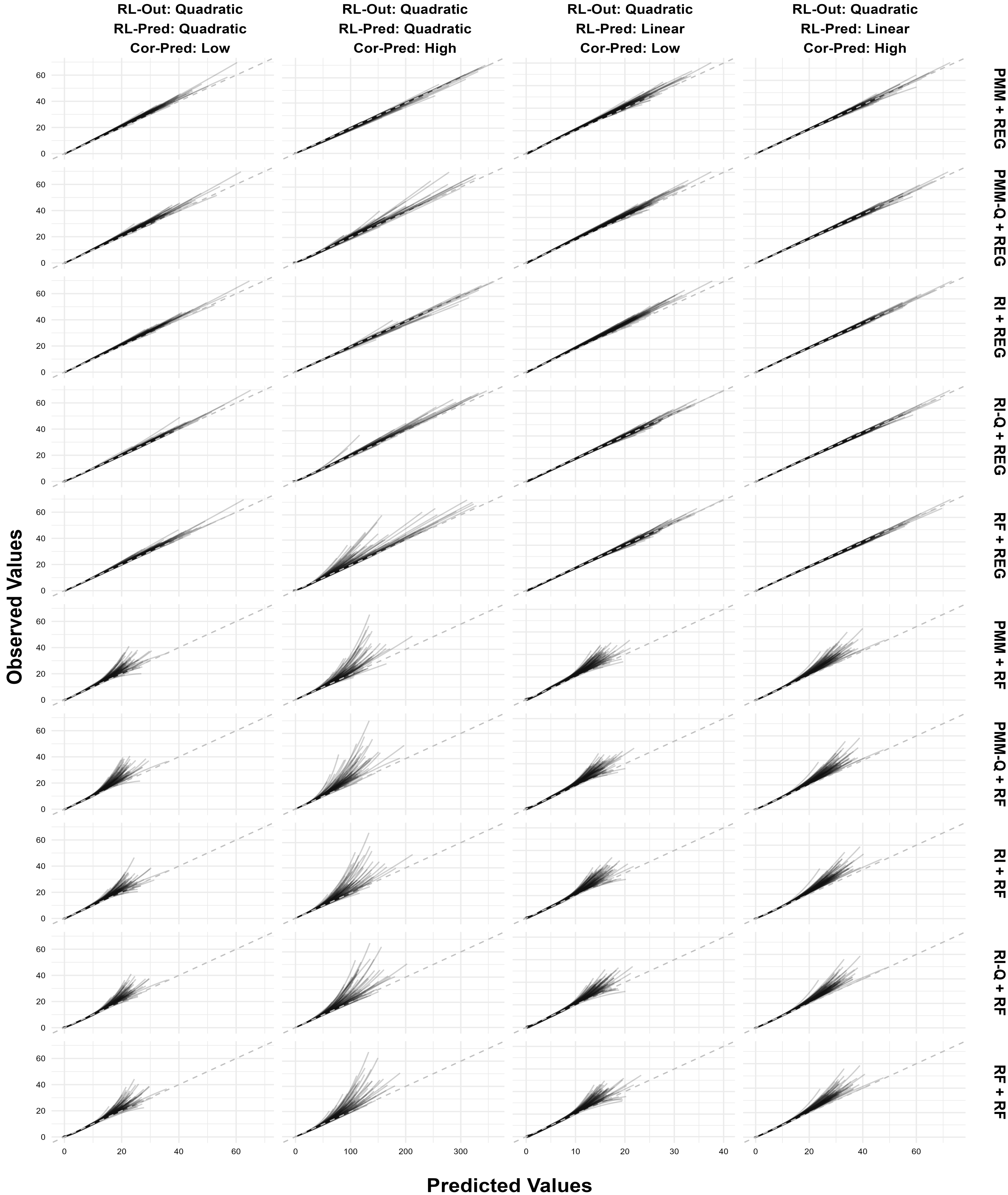

Figure 5: Calibration Plots of different scenarios with a quadratic outcome-relationship (all MCAR).

References

[1] Swj Nijman, A. M. Leeuwenberg, I. Beekers, I. Verkouter, Jjl Jacobs, M. L. Bots, F. W. Asselbergs, Kgm Moons, and Tpa Debray. Missing data is poorly handled and reported in prediction model studies using machine learning: a literature review. *Journal of Clinical Epidemiology*, 142:218–229, 02 2022.

[2] 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.

[3] 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.

[4] 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.

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

[6] 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.

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

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

A Accuracy (RMSE)

Scenario Index	Relationship-Type Between Outcome	Relationship-Type Among Predictors	Relationship-Strength Among Predictors	Missing Mechanism
1	Quadratic	Quadratic	Low	MCmbAR
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	MCmbAR
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	MCmbAR
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	MCmbAR
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	MCmbAR
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	MCmbAR
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	MCmbAR
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	MCmbAR
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 10: (Your caption here)

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 11: Sub-table for RMSE with REG as the prediction model.

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 12: Sub-table for RMSE with RF as the prediction model.

B R^2

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 13: Sub-table for R^2 with REG as the prediction model.

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

Table 14: Sub-table for R^2 with RF as the prediction model.

C Calibration Plots

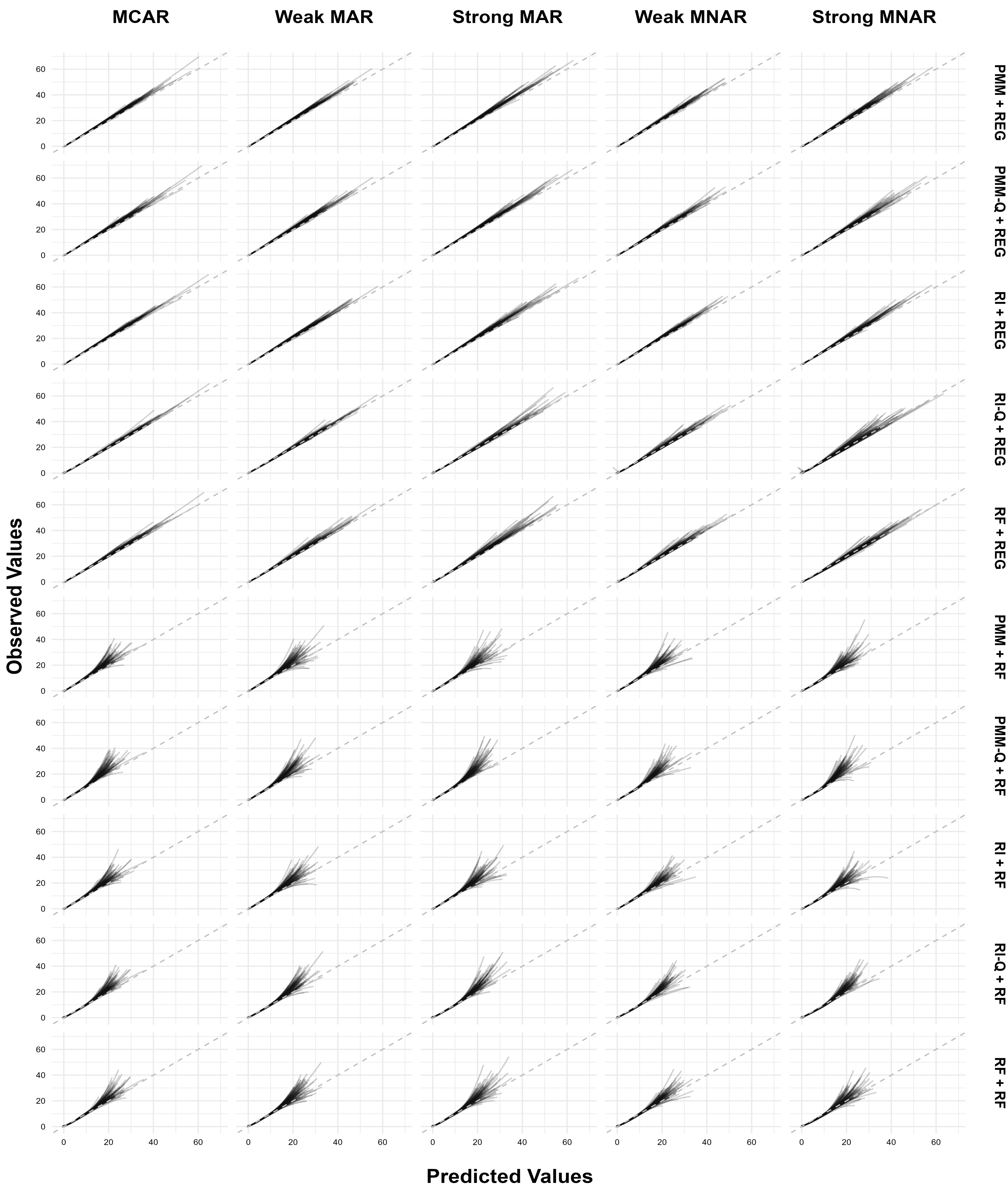

Figure 6: Calibration Plots of RL-Out: Quadratic, RL-Pred: Quadratic and Cor-Pred: Low.

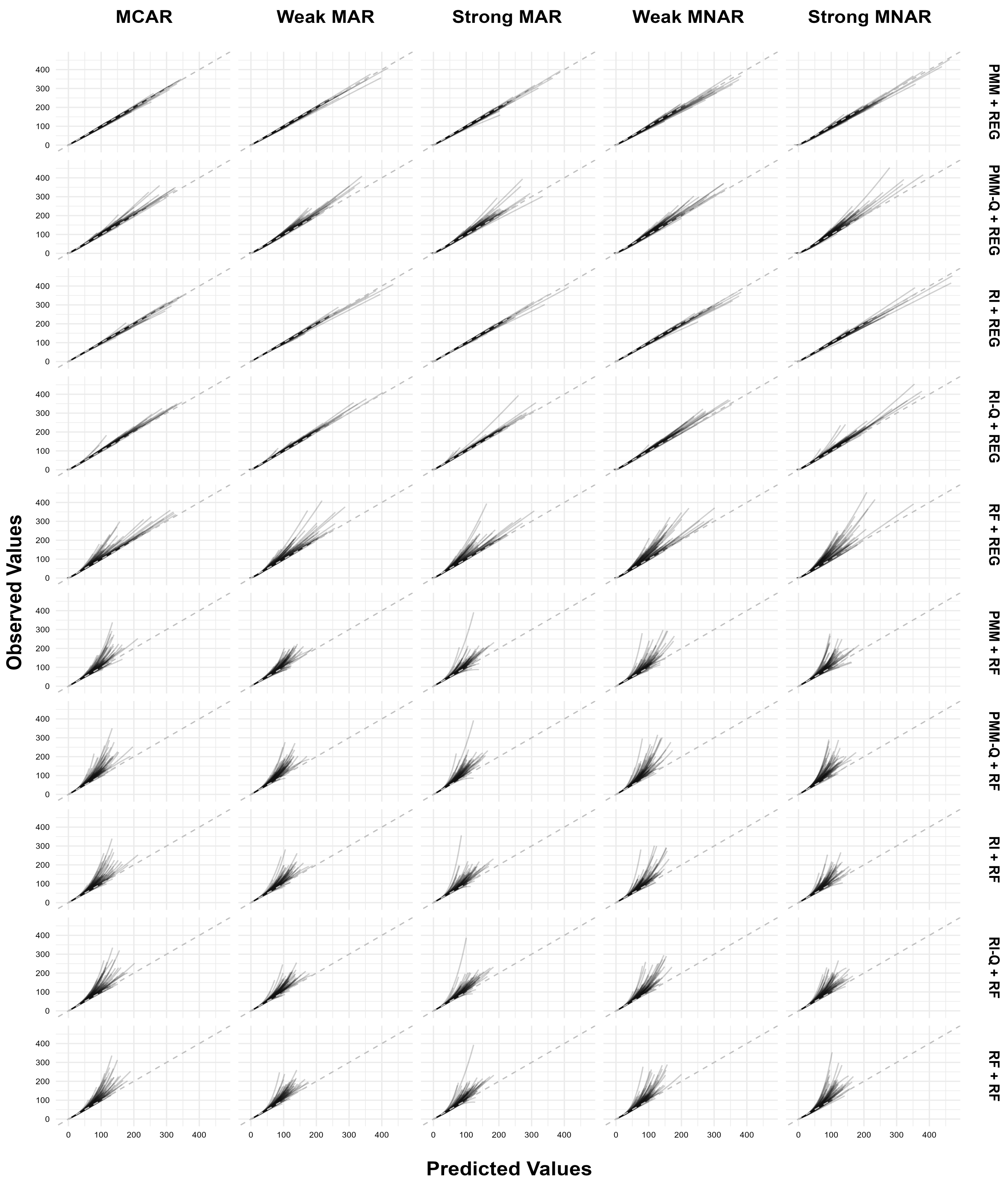

Figure 7: Calibration Plots of RL-Out: Quadratic, RL-Pred: Quadratic and Cor-Pred: High.

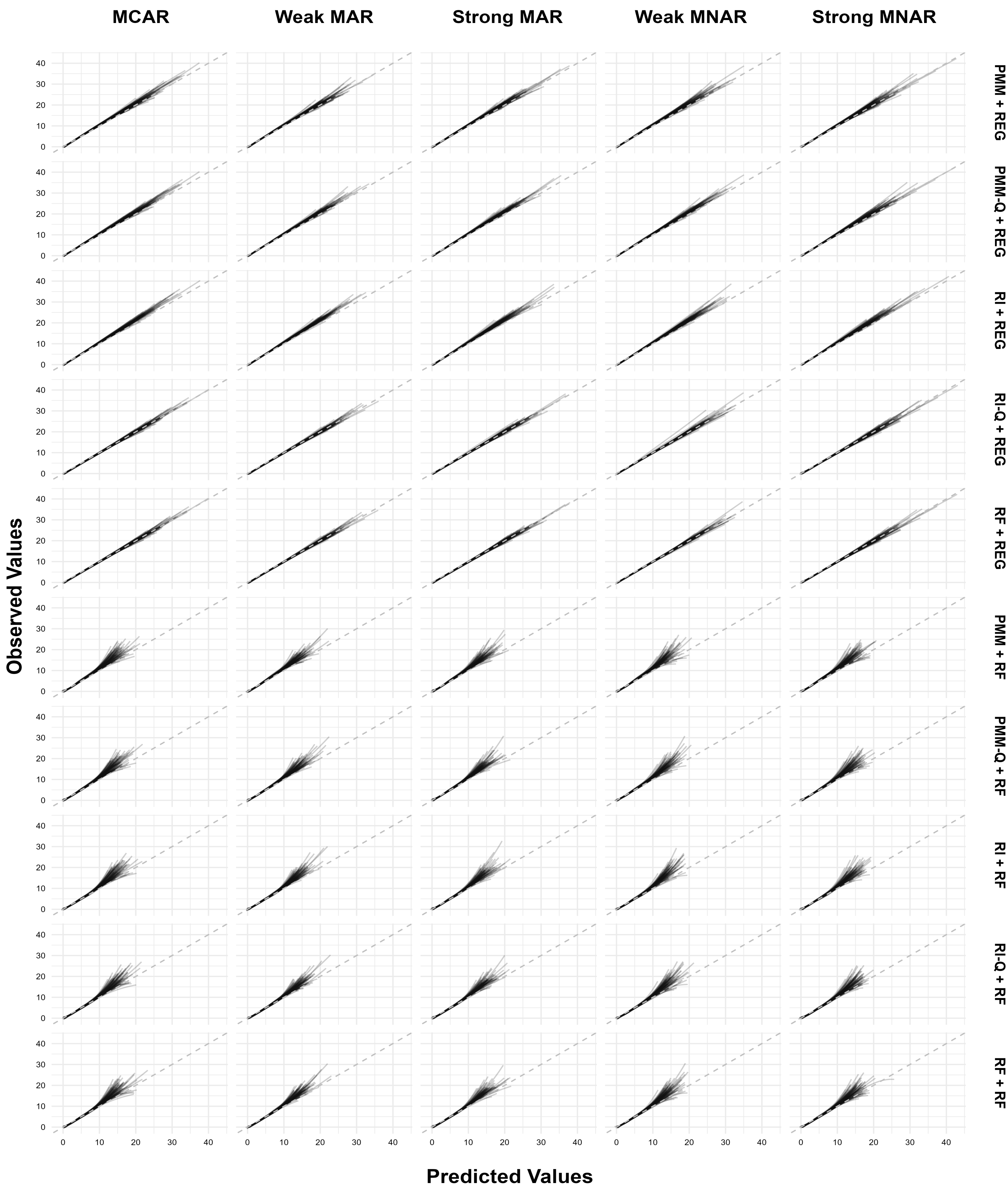

Figure 8: Calibration Plots of RL-Out: Quadratic, RL-Pred: Linear and Cor-Pred: Low.

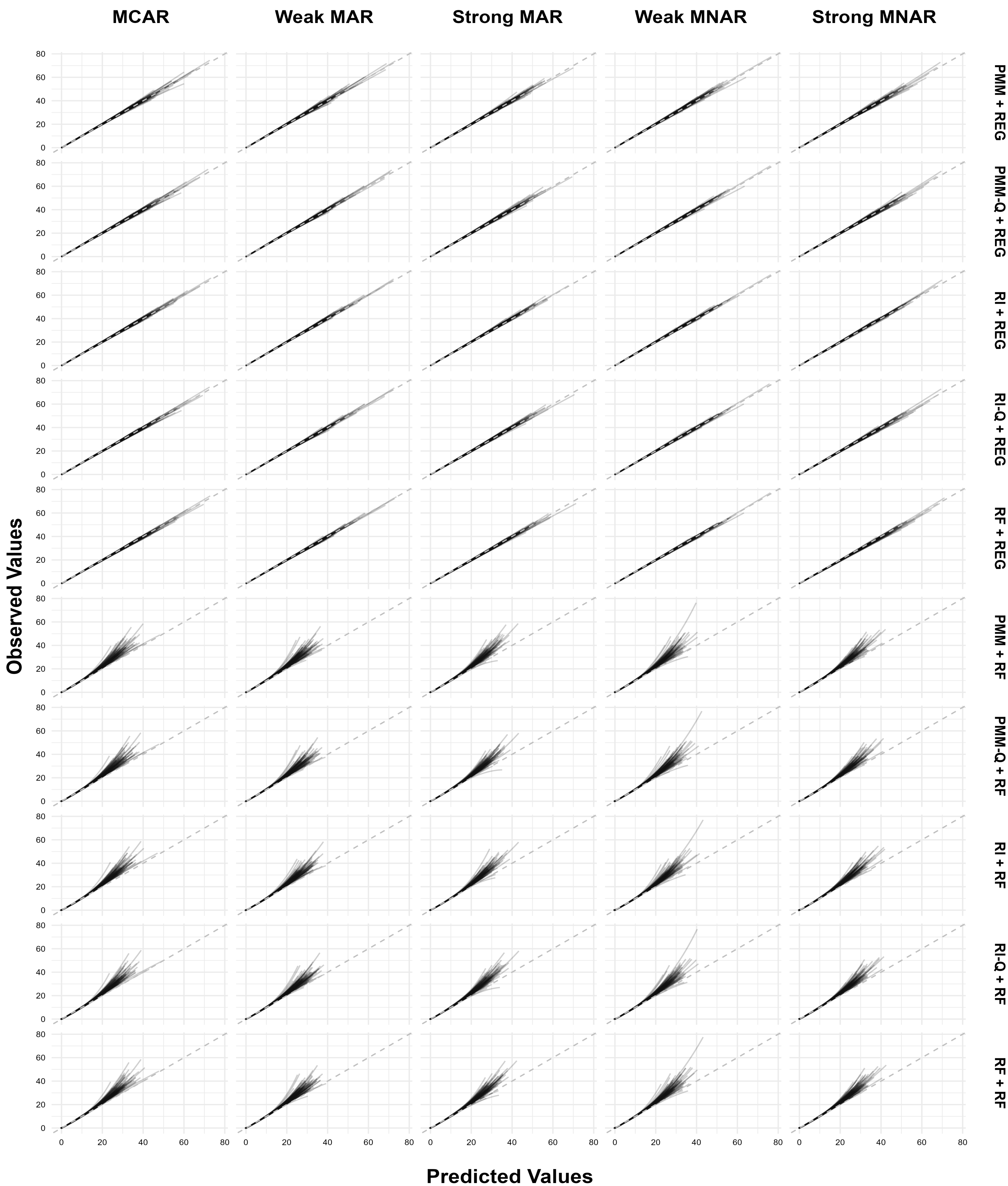

Figure 9: Calibration Plots of RL-Out: Quadratic, RL-Pred: Linear and Cor-Pred: High.

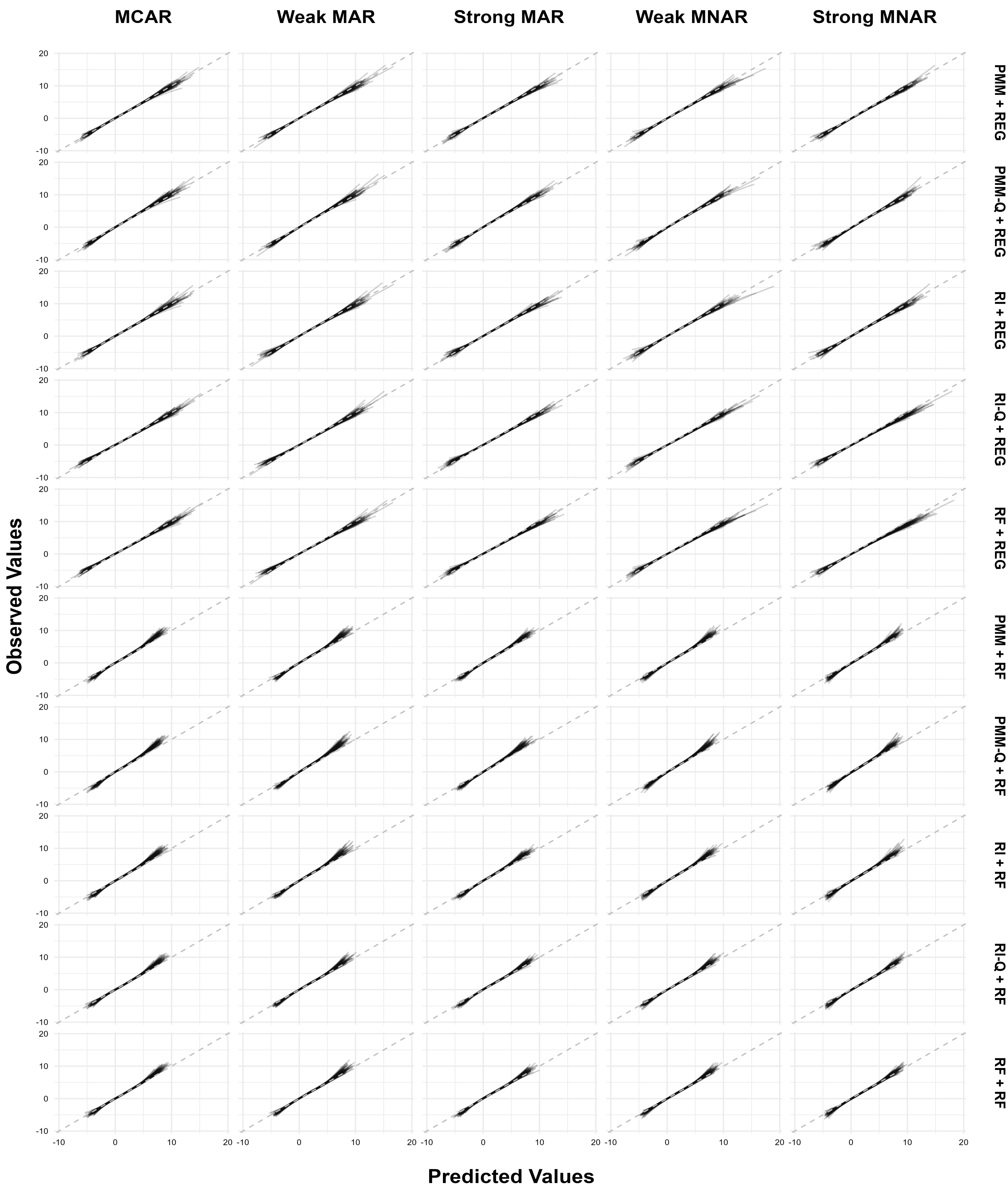

Figure 10: Calibration Plots of RL-Out: Linear, RL-Pred: Quadratic and Cor-Pred: Low.

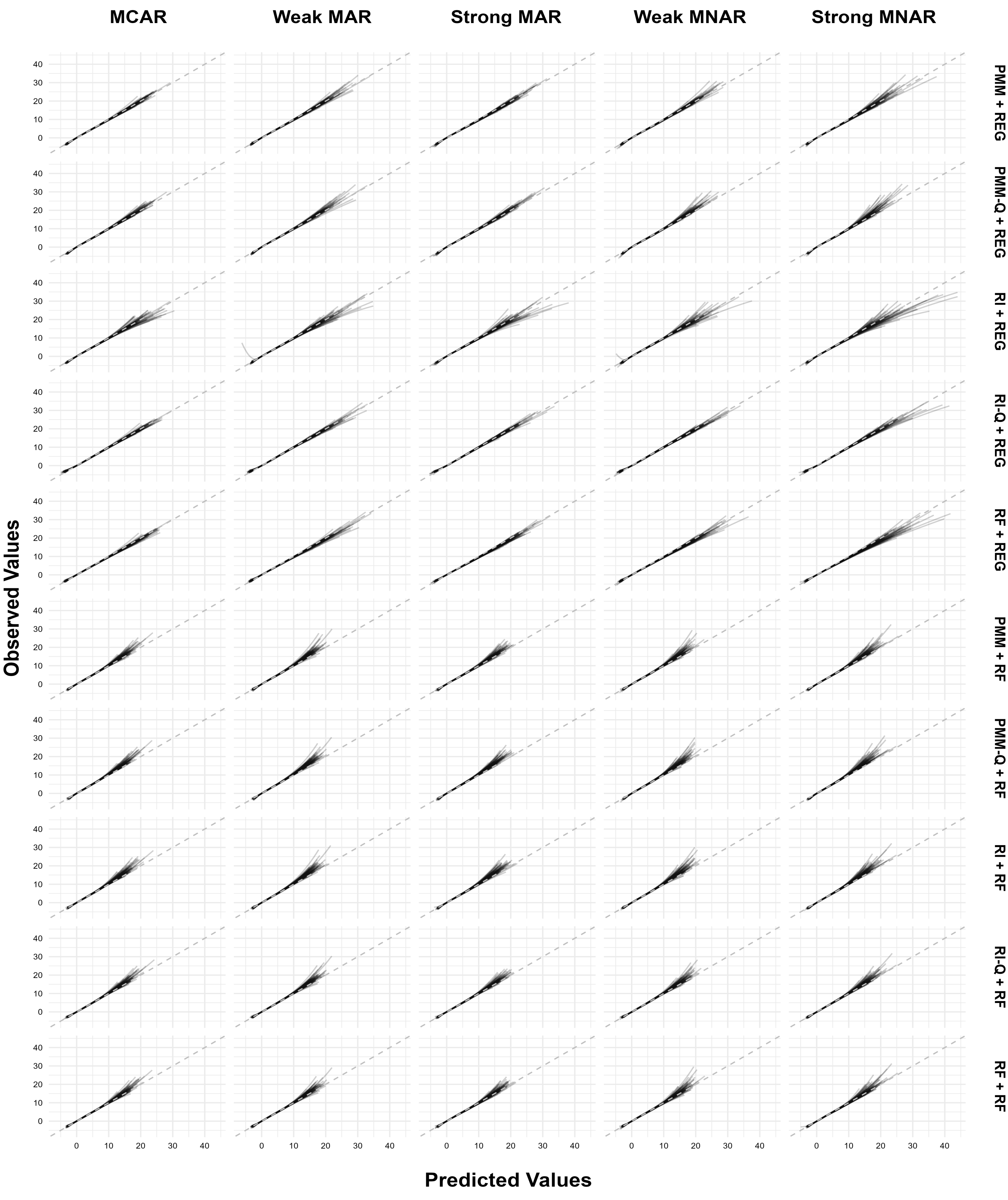

Figure 11: Calibration Plots of RL-Out: Linear, RL-Pred: Quadratic and Cor-Pred: High.

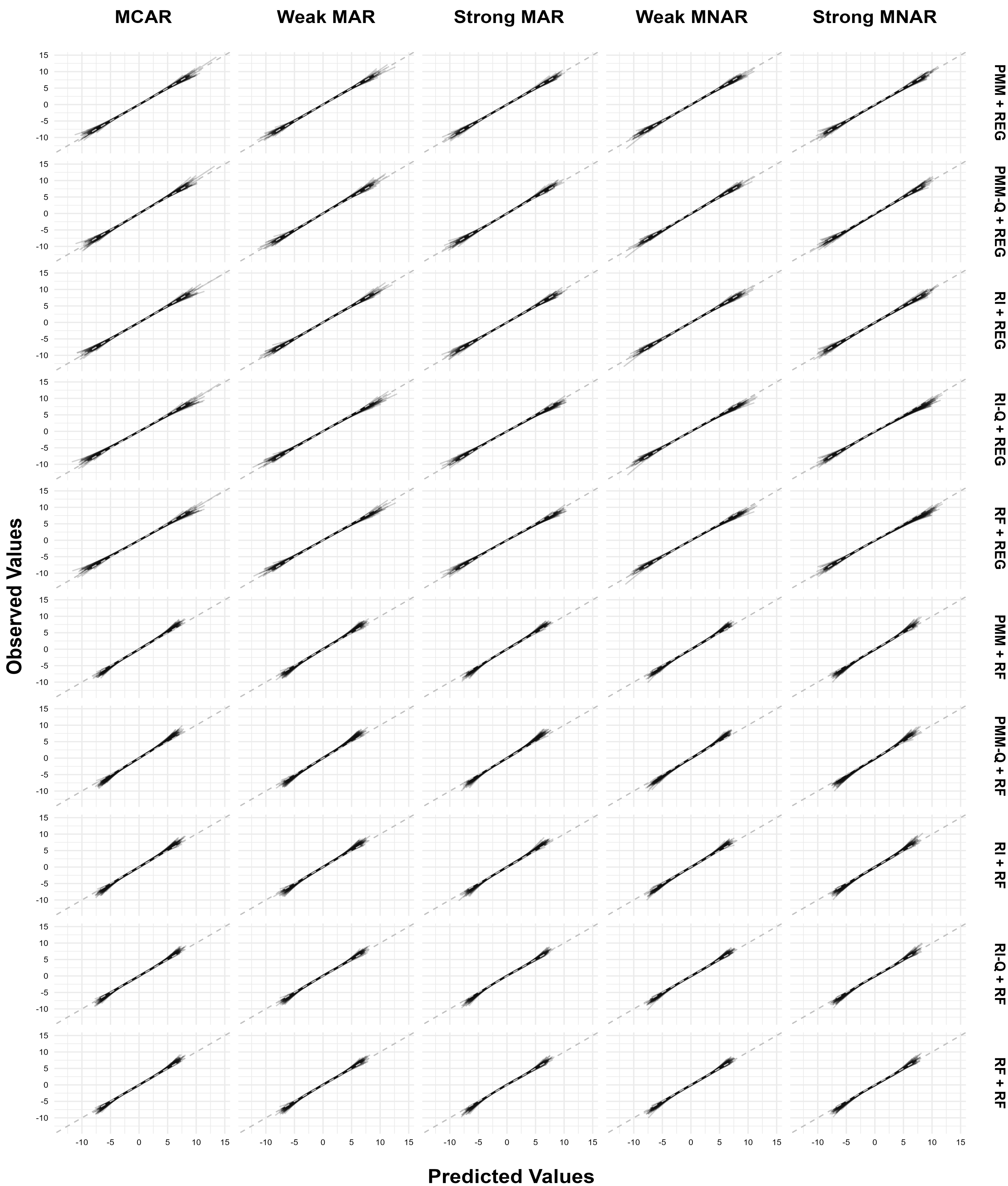

Figure 12: Calibration Plots of RL-Out: Linear, RL-Pred: Linear and Cor-Pred: Low.

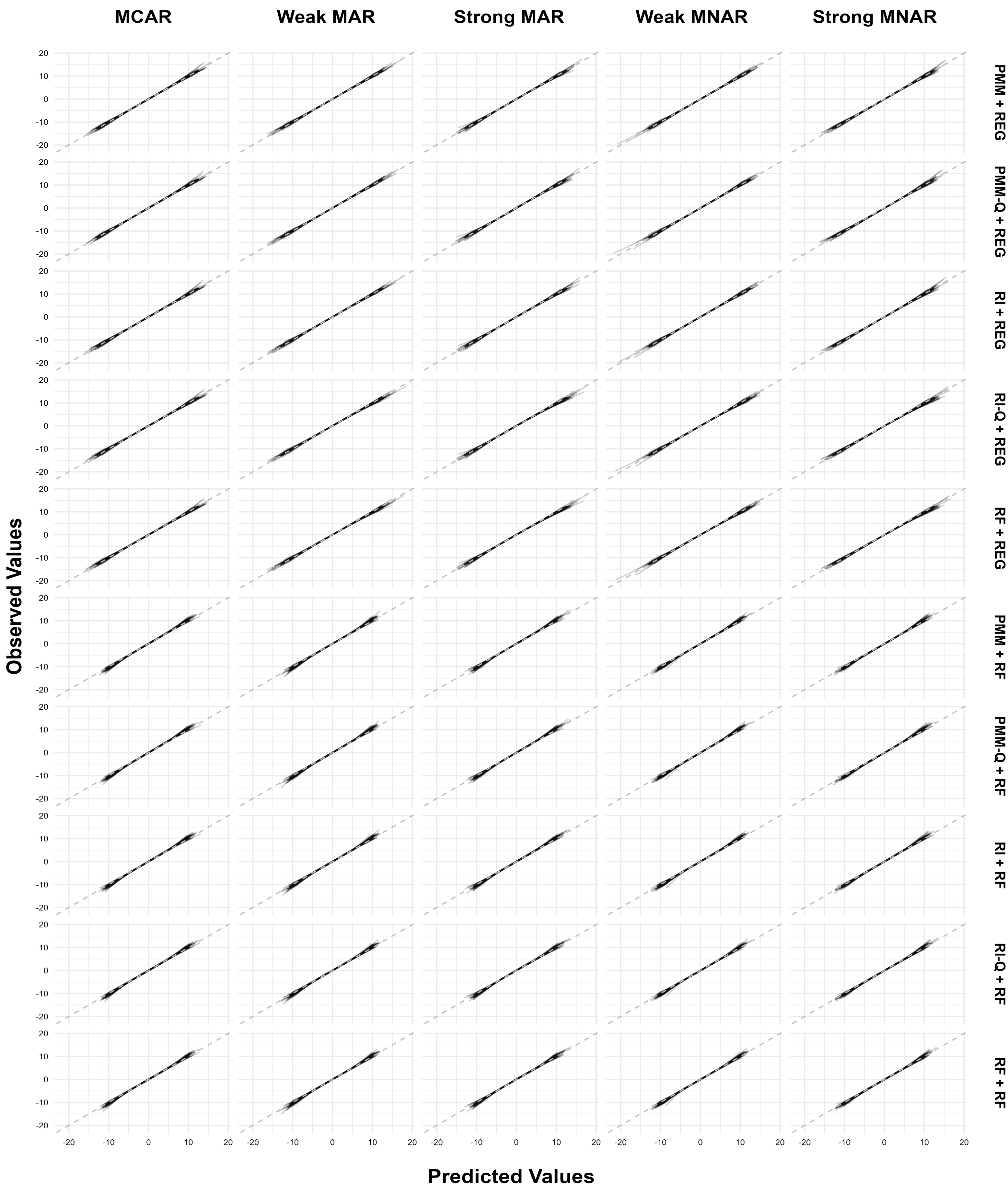

Figure 13: Calibration Plots of RL-Out: Linear, RL-Pred: Linear and Cor-Pred: High.