"Science Stays True Here"
Journal of Mathematics and Statistical Science (ISSN 2411-2518), Vol.6, 109-122 | Science Signpost Publishing

The Effect of Carryover in the Estimation of Model Adjusted Means in Cross-Over Studies

Christos Stylianou
ClinBAY ltd, Cyprus
Email: christos@clinbay.com

Abstract

Carryover effects constitute a potential issue when using the crossover design. Multiple methods exist, both for modelling the carryover effect, and predicting adjusted means from these models. In the current paper, we investigated the reliability of the prediction methods for estimating the model adjusted means and their differences, using a simulation study. Our simulation data suggest that the most reliable estimates were obtained when modelling the carryover effect as a factor, and assuming a carryover of zero in the predictions. However, potential confounding effects can cause large bias when using this method, and in the author's opinion it would be preferable to completely avoid presenting the model adjusted means. Most of the methods tested, as expected, provided identical and reliable estimates of the treatment differences from these model adjusted means, when a carryover was included in the model either as factor, or using a reduced carryover model.

Keywords: Carryover effect, prediction, model adjusted means, LSMEANS, least square means, marginal means.

1. Introduction

Despite being criticized due to a number of potential shortfalls, the crossover design [1] still remains a common method of testing differences among treatments, and is in fact the recommended design for analysis for bioequivalence trials [2], drug abuse studies [3], and in certain circumstances for the thorough QT/QTc studies [4]. One of the most common arguments against the use of crossover designs, is that the effect of the treatment on a subject may be influenced by a previous treatment. In other words, a particular treatment may have what is known as a "carryover effect".

In order to reduce the impact of carryover effects in the design stage of the trial, an adequate washout period is chosen between treatments, to allow for residual effects of previous drugs to be removed from the system of the subjects. This washout period is chosen using available pharmacokinetic data from other studies, but is not necessarily always adequate to completely eliminate all carryover effects. In order to adjust for possible carryover effects, a lagged treatment effect, i.e. treatment in previous periods, may be included in the model, with the first order carryover effect usually being adequate [5].

In clinical trials, model adjusted means are often used as a way to illustrate the treatment effect and treatment differences. These model adjusted means are obtained by predicting from the model, making some

assumptions about the "average" subject. Statistical packages like SAS® and STATA have automatic procedures, LSMEANS and MARGINS respectively, allowing the user to obtain such predictions when a mixed model is used. Although no difference usually exists in the definitions of most model adjusted means, when calculating the difference between treatments due to the carryover effects, different assumptions about the average subject can lead to different estimation of the actual treatment effect. Despite the difference between treatments being the outcome of interest in such trials, it is not unlikely for the adjusted means for each treatment to be used, either for metanalysis, or as priors in later trials.

In this paper we investigated how model adjusted means should be calculated when carryover effects are included in a cross-over trial, using simulations. Additionally, for the adjusted means definitions, which would yield different treatment differences, the performance of alternative definitions was also assessed.

2. Methods

For this paper only the 4x4 crossover Williams design was examined, but results obtained can be generalized to other crossover designs, with the exception of the 2x2 design. The underlying reason for this is that for the 2x2 crossover design, the carryover effect cannot be modelled as a parameter in the model, due to collinearity with the other parameters. As such, carry over effects are usually investigated using an alternative methodology [6]. Additionally, for simplicity purposes, only first level carryover effects were considered for this paper.

Modelling methods

When modelling a crossover trial, a model with sequence, period and treatment as fixed effects, and subject nested in sequence as random effect is commonly considered. If we believe a carryover effect could exist, such a carryover effect is added to the model with one of the following methods:

- Full carryover: Using the value of the previous treatment period (lag treatment), and setting all the Period 1 lag treatments to a non-missing new factor level (usually 0). This technique hence fits to the model, in a trial with "x" treatments "x" fixed effects.
- Reduced carryover [7]: For the observations that have lag treatments that are non-missing (i.e. not Period 1), and not Placebo, create one variable for each with their value being 1 when they occurred, and 0 otherwise. For the lag treatments for Period 1, set all these variables to 0 and when lag treatment is Placebo, set them to -1. This method reduces the degrees of freedom taken by the model, by adding one carryover fixed effect for each treatment except Placebo, with 1 being the treatment occurred at the previous period, -1 if the treatment at the previous period was placebo, and 0 in all other cases. The author notes that although most commonly this method assigns the -1 when lag treatment is Placebo, any of the lag treatments could be selected to be assigned the -1, and this method can still be used when no Placebo treatment is used.

In this paper we considered both models above, and also a model ignoring the carryover completely when calculating the adjusted means.

The author notes that the methodological origin of the reduced carryover model could not be identified. However, to the author's understanding (and experience), the model works by confounding carryover with period, hence shifting the coefficient of carryover and period but not the others. The author hence expected that although the confounding cannot be eliminated when predicting the effects of treatments themselves (hence producing biased adjusted means), it is however limited when predicting the treatment differences as the effects of both period and carryover are subtracted out of the prediction, hence producing valid adjusted means differences.

Model adjusted means methods

When model adjusted means are obtained using a standard package such as least square means (LSMEANS) in SAS® software, the following procedure is applied:

1. Allocate all possible values of the factors of interest (or treatment in this case) one at a time.
2. Assign to the remaining covariates their "average" value as follows:
    a. For continuous variables, the mean value is used.
    b. For class variables, either:
        i. Assume balanced design and assign equal proportions to all categories (this is the default in LSMEANS),

        or

        ii. Check the actual proportions in our sample for each category (this is the observed margins option in LSMEANS).
3. Predict using the specifications above.

As explained above, the default options assume that for the factor variables that are not of interest, equal values between groups are used. This is not an unreasonable assumption in most cases, since we randomize in order to achieve a balanced design, and assume that even if the sample has some differences between the two groups in the population to be predicted, we can assume equal values. However, we need to consider if this assumption is reasonable when a carryover effect is included in the model.

Looking at the literature [8] on how to implement LSMEANS in the SAS® software, we observed that the latter recommended using the 1st method for implementing the carryover effects, using a lag class variable, with the observed margins option. This method, adjusts for the fact that we do not expect equal proportions for each level; which is in fact the case here, since we expect more values to be assigned in the no-treatment group for the carryover variables than the other groups in the 4x4 crossover. However, this method ignores the fact that since this is a crossover design, a treatment cannot have a carryover effect from itself in practice, and hence the latter should be implemented as zero. The options for predicting for both the full carryover and the reduced carryover models above will be identified below, and the potential issues/limitations for each method will be presented.

Full carryover model (abbreviated FM):

a) Marginal means: This is the default option of LSMEANS that would assume a (100/(x+1))% proportion for each treatment (with "x" being the number of treatments). Such a method has the limitation whereby it assumes that an equal number of observations had a carryover effect from each treatment and no-treatment. In fact, in a 4x4 crossover design, one would expect more observations to not have been impacted by a carryover effect, than to have a lag effect from a specific treatment, as all observations in Period 1 can have no carryover effect. This method produces non-estimable results when used in the SAS® software.
b) Marginal means, observed margins: Although similar in spirit to (a) this method takes into account that not all carryover treatments were observed equally, as not all observations in Period 1 have no carryover effect. As mentioned above, this appears to be the method suggested by the SAS® software for predicting when a carryover is included in model [8].
c) No carryover: This method works by assuming in the predictions that no carryover effects exist (i.e. all carryover effects will be set to zero). This method appears to make a reasonable assumption that in the population under consideration, we are not interested in any carryover effect, but only in the actual drug effect. However, even though that is a reasonable assumption, the problem appears in a theoretical level on our models, if we can assume that no confounding occurs between the treatment variables and carryover variables. This method produces non-estimable results when used in the SAS® software.
d) Mean carryover: For each treatment, the observed proportions of the lagged treatment should be calculated and used in the estimates command. This approach may make sense, if we believe that confounding exists and the model estimates are biased because of confounding. If the model is estimated correctly, this method has a larger potential for bias as different carryovers are added in each group. Another possible downside of this method is the increase in complexity as treatments/periods increase.

The treatment differences of options a, b and c, are the same as the effects from carryover are eliminated when subtracting the treatments, despite providing different adjusted means. Only option d will yield different model adjusted mean treatment differences estimates, as it has different carryover effects for each treatment.

The reason the SAS® software provides non-estimable results is not explicitly stated, and the most likely reason out of the reasons provided based on SAS® software documentation [9] for this, is that "There is a confounding problem that is associated with the data and the model". The author's understanding of this error, is that this is meant as a safeguard to avoid predictions where it should not predict. However, such automated checks are prone to errors, as they cannot fully conceptualize the study design (as a design issue would probably lead to non-estimable coefficients rather than predictions, and may not be applicable in the scenario described here). It has to be noted that other packages like STATA do not forbid such a prediction.

Reduced carryover model (abbreviated RM):

a) No carryover: Similarly to the FM, this will plug in the model that no carryover effects exist.
b) Marginal means: As the carryover effect is from a continuous variable, this will calculate the mean of each carryover variable and plug it in the model.

The two options for the RM will be the same when no missing data exist (as the mean of all carry over variables will be zero), and due to the low numbers of missing data, little difference was expected between the two. Despite providing different adjusted means, their treatment differences are the same as the effects from carryover are eliminated when subtracting the treatments.

No carryover model (abbreviated NCM)

For completeness, a model ignoring the carryover effects completely was also implemented. The standard marginal means were used for the predictions from this model.

Simulation methods

The aim of the simulations was to assess the performance of the model adjusted means methods and their differences, in terms of prediction accuracy. Their performance was assessed using the mean bias, maximum absolute bias and MSE (Mean Squared Error). The simulation was performed in STATA 11 [8], and the modelling procedure used was xtmixed.

Simulation design

A 4x4 crossover study was conceptualized based on author's experience. Results were simulated from a model with an intercept of 1.5 units, a "0 units" effect of Period 1, a "-1 units" effect of Period 2, "-0.5 units" of Period 3, "-0.25" units of Period 4, "5 units" effect of Treatment 1, "4.5 units" effect of Treatment 2, "3 units" effect of Treatment 3, and "1 unit" effect of Treatment 4. Random deviation was added from a Normal distribution; a 0.2 standard deviation was added at a subject level and 0.5 at the observation level.

Sample size calculation

The sample size in each simulation was calculated based on STATA 11 built-in sample size calculator [10], assuming the study wanted to detect a difference of 0.5, with 80% power and alpha=0.05 and needed 19 subjects. As it is standard practice in clinical trials, a larger number was chosen based on a predicted missingness, and 20 subjects were used in each simulation.

Number of replications

The outcome of these simulations was the prediction bias of these methods; the accuracy to detect a 0.01 bias was considered to be sufficient. Based on a test simulation design without a carryover effect, the maximum standard deviation detected for bias was 0.2. To estimate the coefficients of the model with 95% confidence, the number of replications needed as per [11] was calculated to be 1536.64. The number of replications for each Scenario at each carryover level was hence set to 1550.

Scenarios tested in the simulation

The following 3 scenarios were examined individually in the simulation:

- Scenario 1: Only one treatment had a carryover effect.
- Scenario 2: Three of the four treatments had a carryover effect. This could be the case the trial includes a Placebo effect (which does not have a real effect to carryover).
- Scenario 3: All four treatments had a carryover effect. This could be the case when all four treatments comprise active compounds, or the Placebo effect has a carry-over (such a "carry-over" could occur in a case of questionnaire data where a subject overestimated the effectiveness of the drug in the next treatment because of the lack of effect of placebo in the previous treatment).

Each scenario above was investigated with and without missing data. For the scenarios with missing data, the data were set to missing under a missing completely at random assumption, with the relatively high proportion of 10% missingness. The reason missing data were investigated separately, is because some of the prediction methods above relied on the observed proportions, and could hence be affected by the imbalance caused by the missing data.

For all scenarios above, the carryover effect was investigated as a percentage of its respective lag treatment effect, from 0 to 10% (tested at each 1% increment) of the real treatment effect at the last period.

3. Results

The MSE was considered the primary measure of the methods' performance, with the results of the adjusted means shown in Figure 1. Assuming no carryover effect in the FM had a relatively flat MSE, which suggests a relative stability in its estimates, irrespective of the carryover effect and the Scenario. The marginal means and the marginal means observed margins of the FM model yielded very similar results in terms of MSE in the simulation, and both showed an increasing trend as carryover increased. As expected, these methods performed relatively better in Scenario 1, compared to Scenarios 2 and 3. Non-surprisingly, both methods of the RM produced almost identical results in terms of MSE, even in the case with missing data. These methods also showed an increasing trend, but they gave slightly lower MSE estimates when compared to the FM marginal means methods. The performance of the no carryover model was heavily reliant on the Scenario tested and the level of carryover effect. This technique outperformed the others in terms of MSE in cases with very low carryover.

The mean bias results are shown in Figure 2. The FM with no carryover effect had a relatively flat mean bias of 0 in Scenarios 1 and 2 (with the exception of the Scenario 2 with missing for treatment 4). Even in the cases where the bias was not flat and instead showed an increasing trend, this method demonstrated a lower bias compared to the other methods. All other methods showed an increasing relationship with carryover proportion in terms of mean bias, with their performance varying between each scenario.

The maximum absolute bias results are shown in Figure 3. The FM with no carryover effect had a relatively flat maximum absolute bias in most scenarios, but the maximum bias varied between scenarios and with the presence or absence of missing data. In Scenarios 2 and 3, and with high carryover, this method

appeared to have outperformed the others, especially when no missing data existed. All other methods showed an increasing relationship with carryover proportion in terms of maximum absolute bias, with their performance varying between each scenario.

MSE estimates of the adjusted mean difference are shown in Figure 4. Both with and without missing data, the MSE of all the FM methods except the one with mean carryover and RM was flat, indicating that it was not impacted by the carryover effect. When no missing data existed, both the FM with mean carryover and the NCM showed an almost identical increasing relationship, with the MSE of the bias varying depending on the Scenario. When missing data existed, the FM with mean carryover exhibited a greater MSE to all other methods, and the NCM exhibited a similar behavior as with the no missing data scenarios.

Similar relationships to the ones with MSE were observed for the mean bias and the maximum absolute bias, with FM methods except the one with mean carryover and RM having flat relationships with carryover for both measures, and the other two methods varying depending on the carryover. The results of the mean bias and maximum absolute bias for the treatment difference are shown in Figures 5 and 6 respectively. Most notably the average bias was zero for the FM methods, except the one with mean carryover and RM.

4. Discussion

The simulations have shown that the best model adjusted mean predictions were obtained in cases with high carryover effects, when the model using the lag treatment as fixed effect was used, assuming no carryover effect in the prediction. Other prediction methods from the same model, the model with reduced carryover, and the no carryover model, varied in terms of their performance depending on the Scenario, and were unsurprisingly more prone to error as the carryover effect increased. Based on the results of the simulation, the author believes that in cases of high carryover effect and sufficient sample size, this method has the highest chance to yield the best predictions. The limitation of the SAS® software to provide this prediction is not unreasonable, as this prediction showed a relatively flat maximum bias, which suggested that the confounding concern should not be taken lightly. However, further simulation work may be needed to investigate other scenarios, and especially cases with larger sample size, as the confounding problem could be reduced as sample size increases.

The author believes that in light of this simulation work, presentation of model adjusted means should ideally be avoided, when there is reason to believe that carryover effect exists, or the reduced carryover effect method is used. Unlike in other contexts, the FM out of the package adjusted means estimates would probably yield a result with bias being quantifiable (quantified based on the model coefficients), and if presented, the bias should be acknowledged. The author finally notes that the best prediction method would be from the FM, assuming no carryover in the prediction, but its results should be presented with caution.

When predicting the treatment differences of model adjusted means, all the FM methods except the one with mean carryover and RM methods, provided stable estimates and appeared to be unbiased on average. Despite RM providing biased predictions of the model adjusted means themselves, the estimates of the differences were identical for all Scenarios tested here to the ones of the FM as expected. Hence, the RM can be used to save degrees of freedom, if the model adjusted means themselves are not to be presented.

In conclusion, presentation of model adjusted means should be avoided when there is reason to believe carryover effects exist. If adjusted means are needed, but a carryover effect is present, the FM, assuming no carryover in the prediction, has the highest likelihood of providing valid estimates, but the potential bias should be acknowledged. Treatment differences of model adjusted means can be obtained with accuracy, for both the FM and the RM, as long as the adjusted means themselves used the same average values for the covariates between treatments in their predictions.

References

[1] P. M. Simpson and R. M. Hamer, "Cross Crossover Studies Off Your List," *SAS Users Gr. Int. Annu. Conf. 24th, SAS Users Gr. Int.*, pp. 1303–1309, 1999.

[2] Fda, "Guidance for Industry: Establishing Bioequivalence," no. January, 2001.

[3] U. S. D. of H. and H. Services and F. and D. Administration, "Guidance for Industry Assessment of Abuse Potential of Drugs Guidance for Industry Assessment of Abuse Potential of Drugs," *Draft Guid.*, no. January, pp. 1–25, 2010.

[4] Ich E14, "Guidance for Industry Interval Prolongation and Guidance for Industry," *U.S. Dep. Heal. Hum. Serv. Food Drug Adm. Cent. Drug Eval. Res. Cent. Biol. Eval. Res.*, no. October, 2005.

[5] C. W. G., *Experimental Designs*. Asia publishing house, 1957.

[6] S.-C. Chow and J. Liu, *Design and Analysis of Bioavailability and Bioequivalence Studies*, 2nd ed. New York: Marcel Dekker, Inc, 2000.

[7] R. J. Littell, R. C., Stroup, W. W., & Freund, *SAS for linear models*. SAS Institute, 2002.

[8] S. Institute, "Example 65.7 Crossover Designs," *SAS/STAT(R) 9.2 User's Guide*. .

[9] S. Institute, "Usage Note 22582: Nonestimable result from CONTRAST, ESTIMATE, or LSMEANS statement," *KNOWLEDGE BASE / SAMPLES & SAS NOTES*. .

[10] StataCorp, "Stata Statistical Software: Release 11." College Station, TX: StataCorp LP, 2009.

[11] A. Burton, D. . Altman, P. Royston, and R. . Holder, "The design of simulation studies in medical statistics," *Stat. Med.*, vol. 25, pp. 4279–4292, 2006.

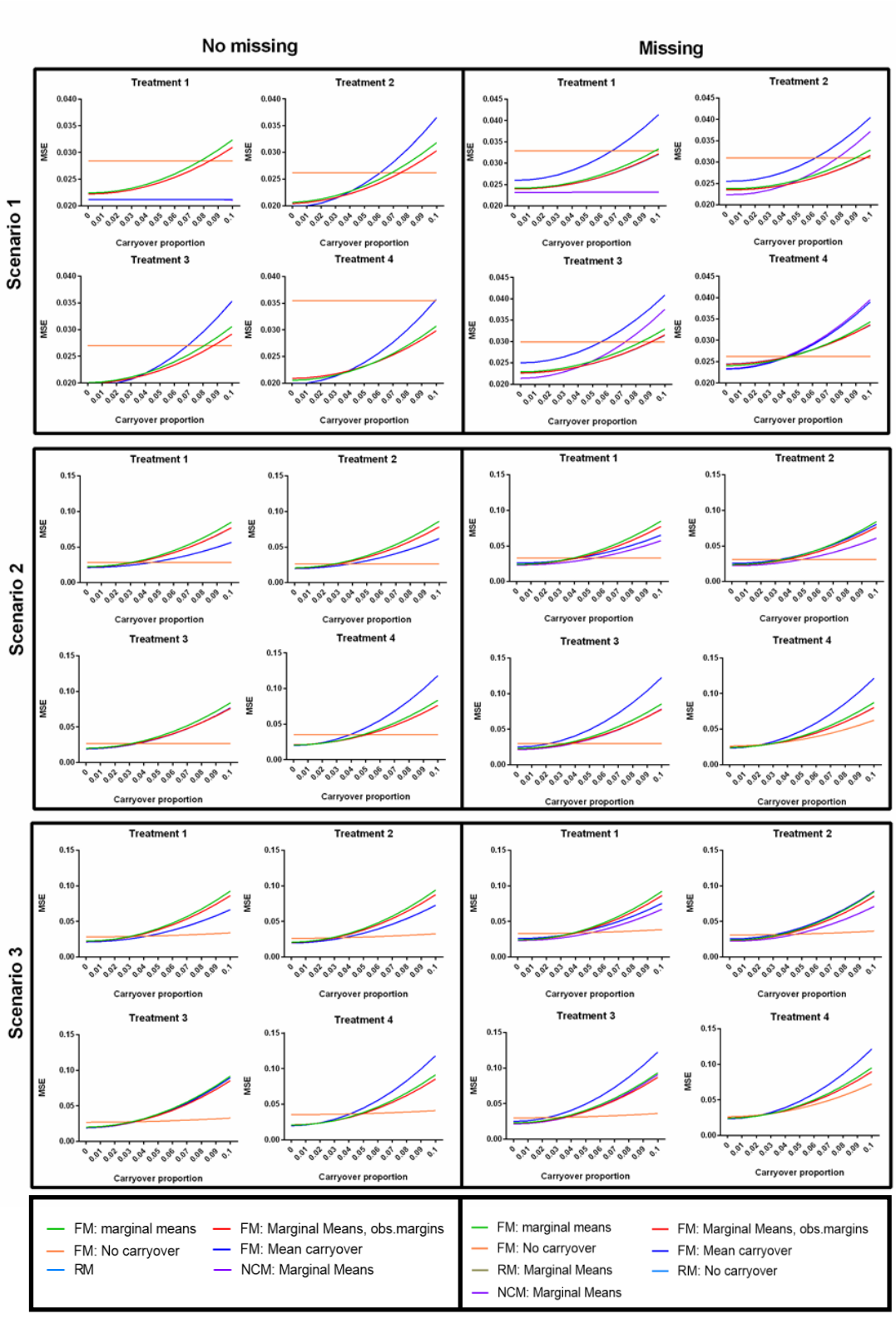

Figure 1: MSE of adjusted means per treatment as a function of carryover proportion for each Scenario, with and without missing data.

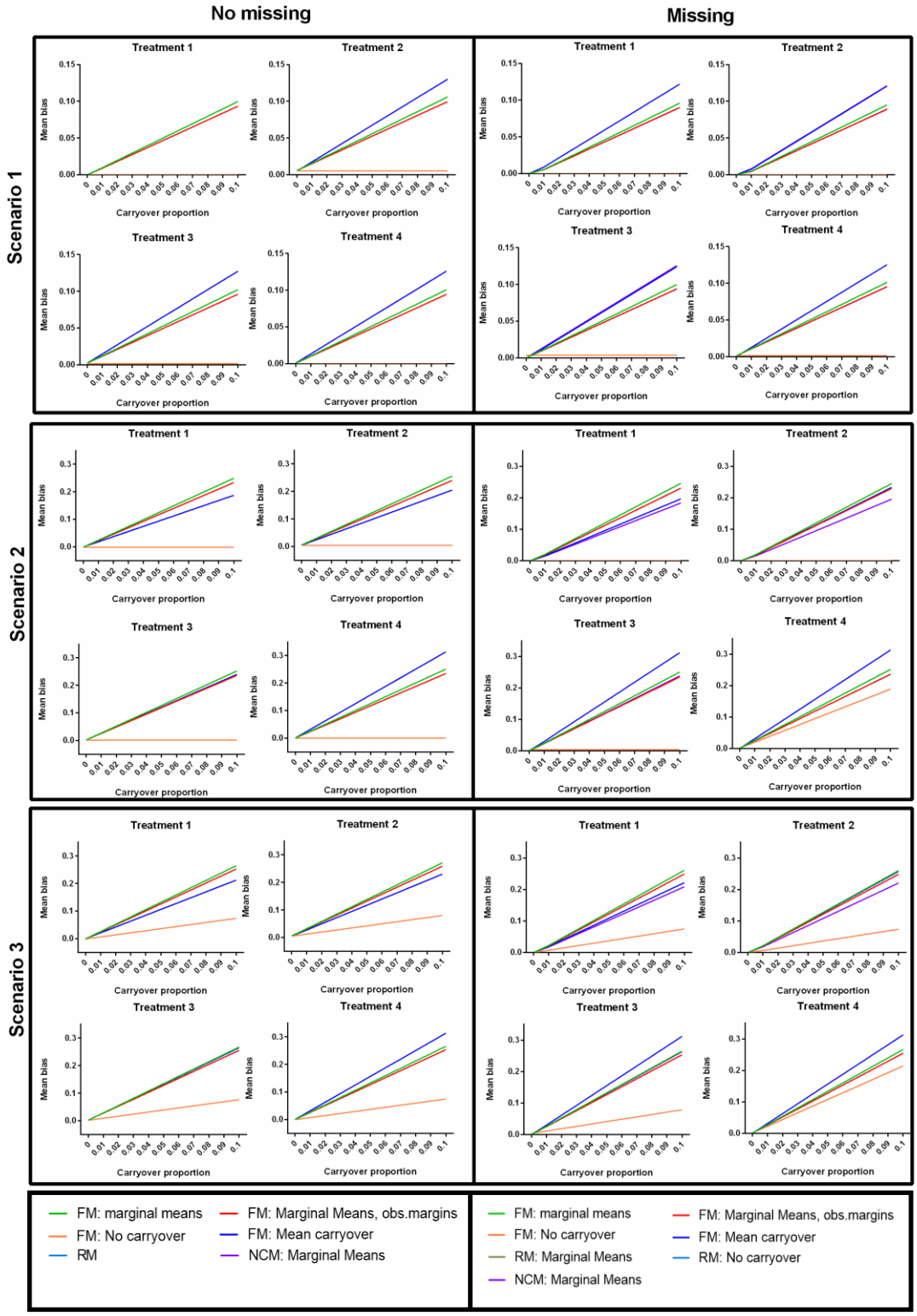

Figure 2: Mean bias of adjusted means per treatment as a function of carryover proportion for each Scenario, with and without missing data.

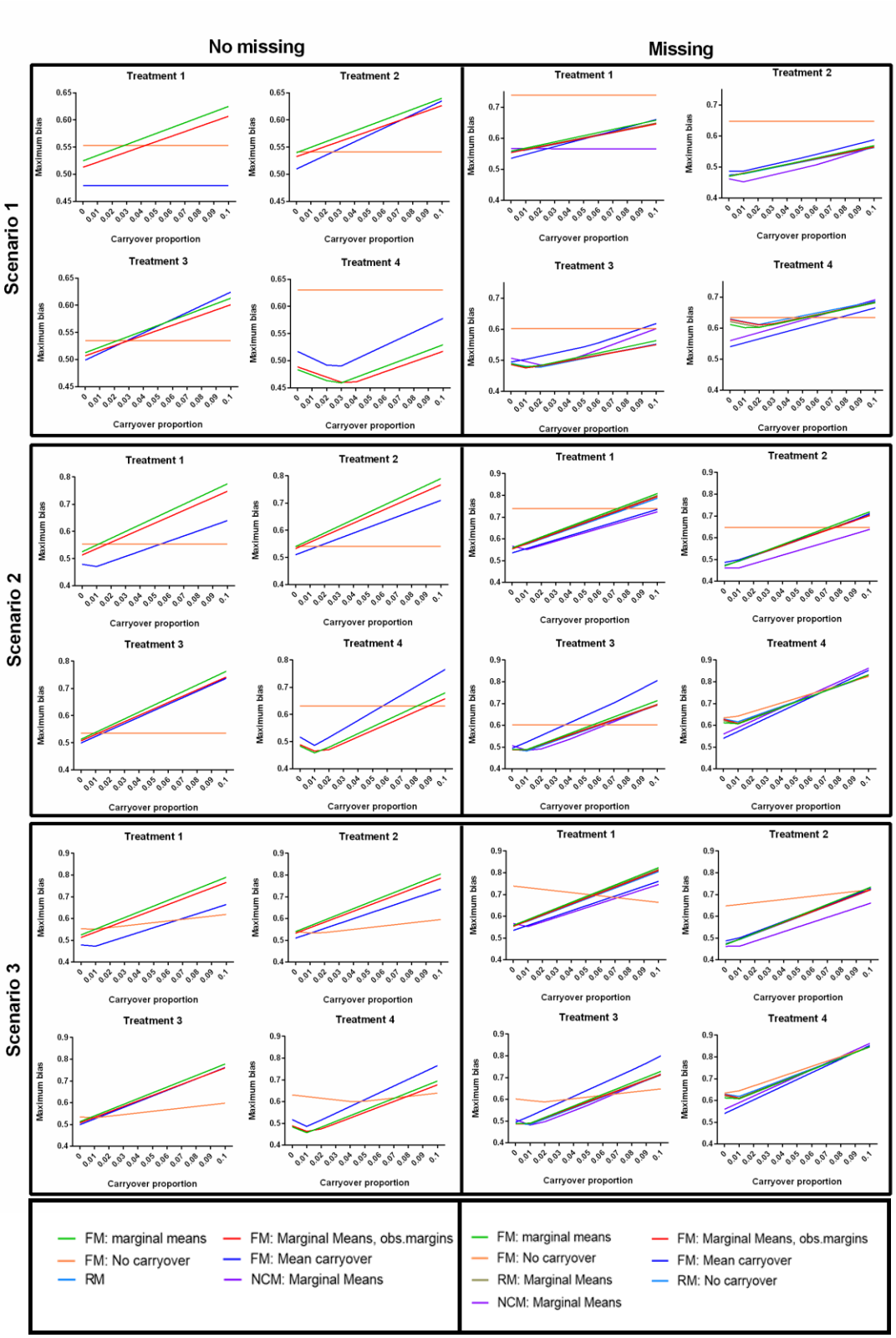

Figure 3: Maximum absolute bias of adjusted means per treatment as a function of carryover proportion for each Scenario, with and without missing data.

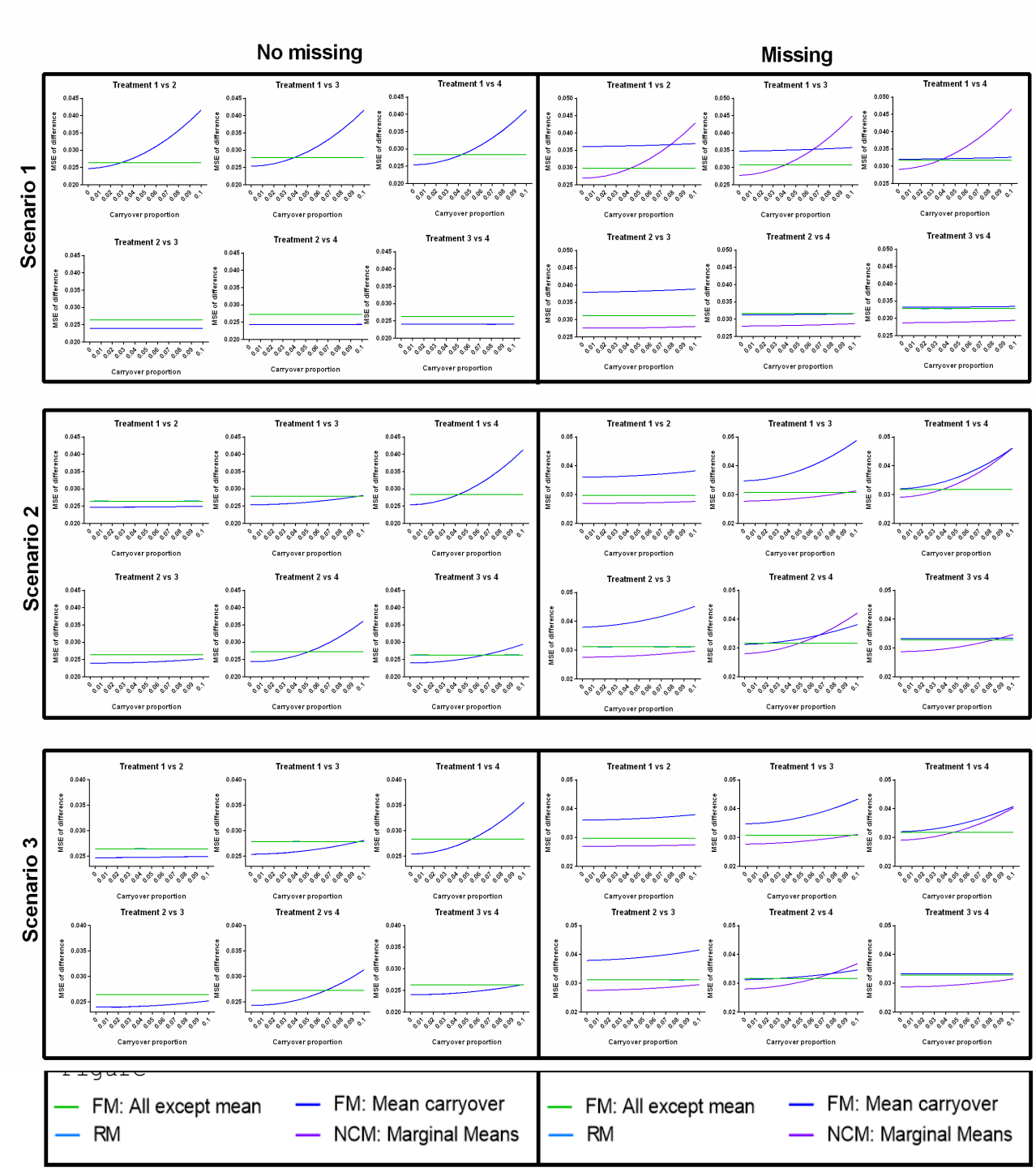

Figure 4: MSE of treatment adjusted means differences as a function of carryover proportion for each Scenario, with and without missing data.

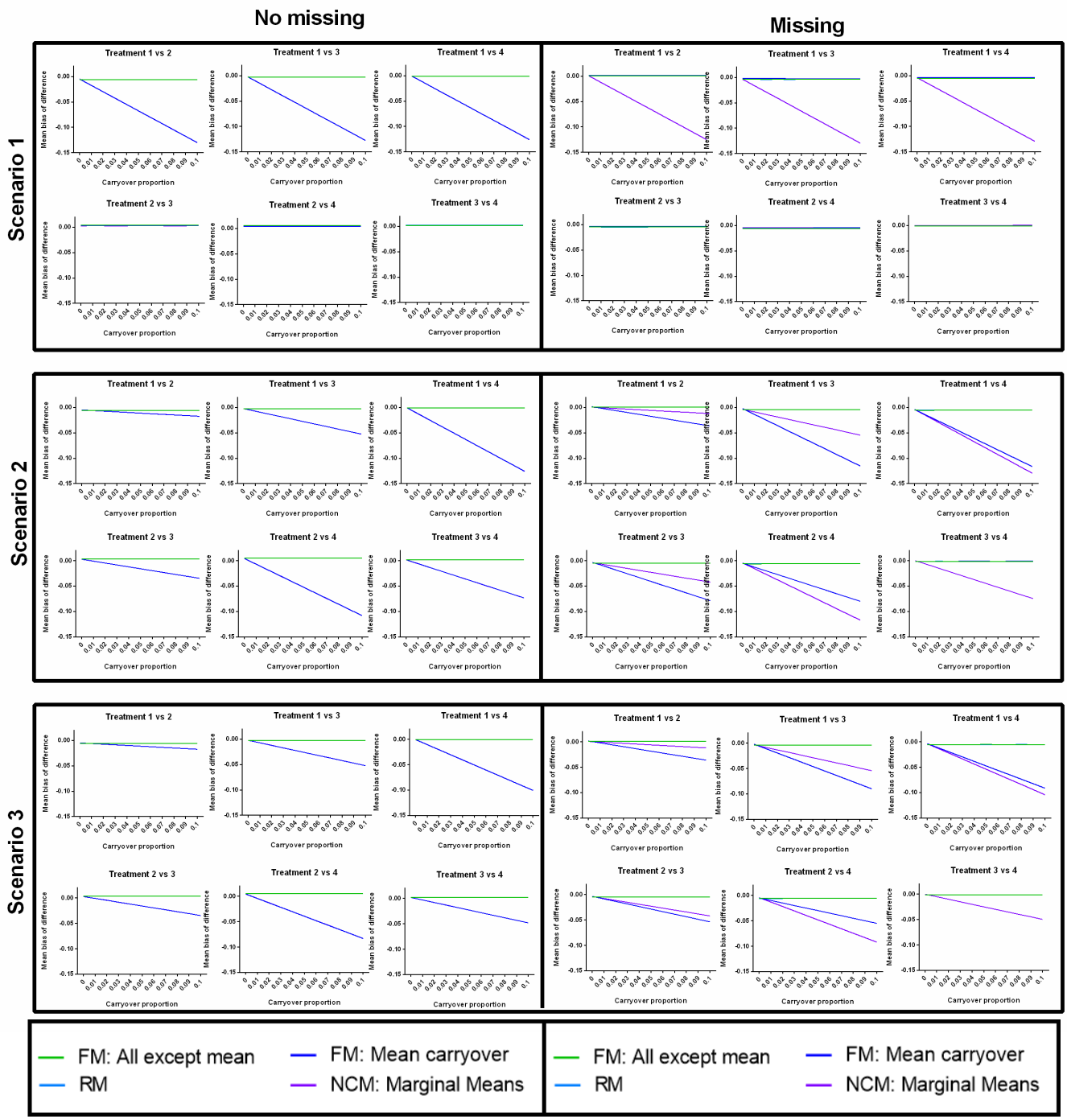

Figure 5: Mean bias of treatment adjusted means differences as a function of carryover proportion for each Scenario, with and without missing data.

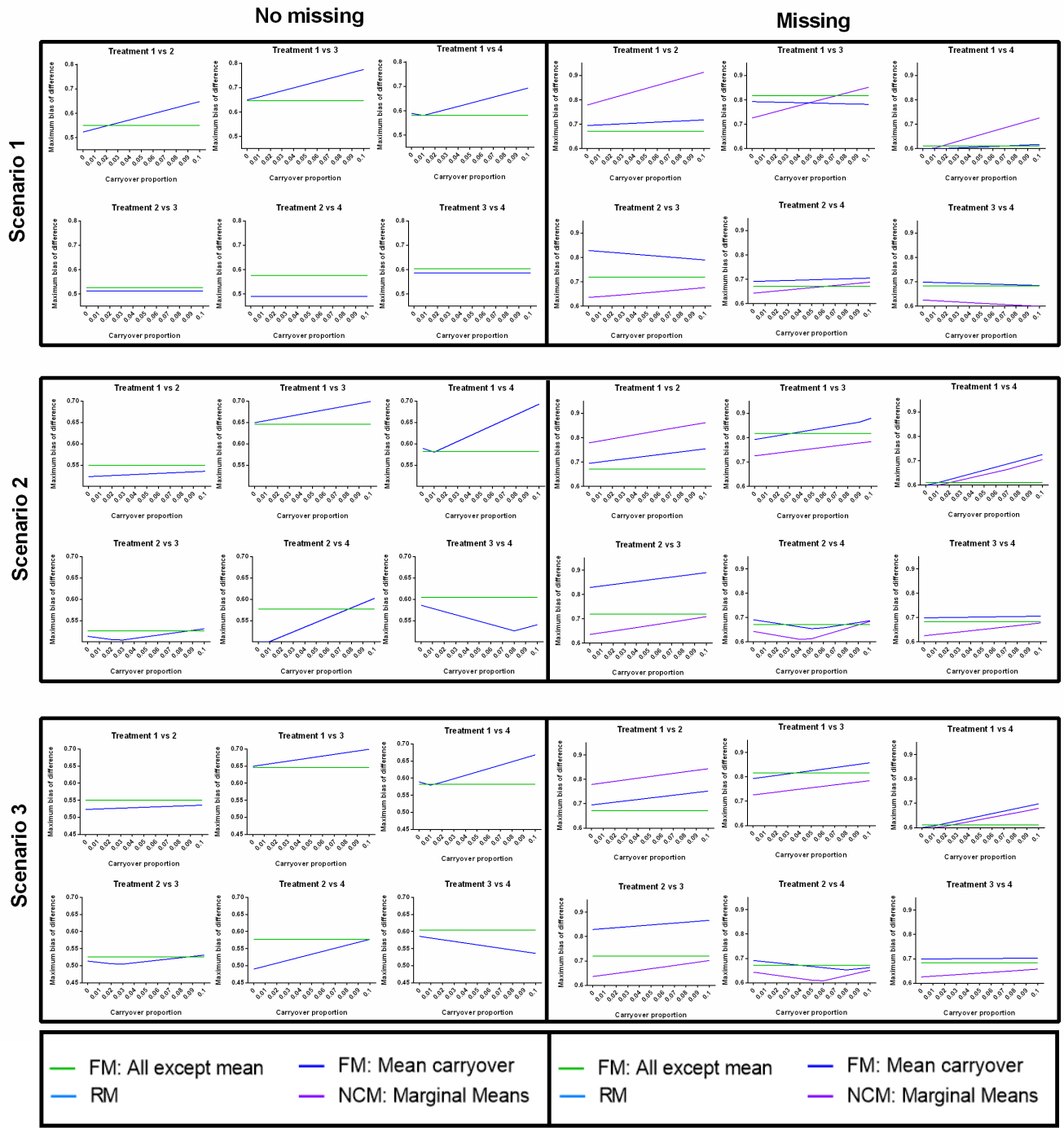

Figure 6: Maximum absolute bias of treatment adjusted means differences as a function of carryover proportion for each Scenario, with and without missing data.