Multiple Imputation Strategies for Missing Continuous Outcomes in Non-Inferiority Randomized Trials

Juan Abellan[1], Andre Acusta[2], Lin Taft[2], Douglas Thompson[3]

[1] GSK, Advanced Biostatistics and Data Analytics, UK; [2] GSK, Quantitative Sciences Clinical Statistics, US; [3] GSK, Quantitative Sciences Clinical Statistics, UK

Background

In contrast to superiority trials, the consideration of how to appropriately handle missing data in non-inferiority (NI) trials has received much less attention in the research literature. NI trials require the same level of attention when defining an estimand, however, reference-based methods (e.g. jump-to-reference, which are frequently conservative in the case of superiority trials) are not appropriate in NI trials due to the direction of bias making it more likely that the conclusion of non-inferiority is met.

Therefore, the motivation for this work was to consider what would be an appropriate reference from which to borrow information in NI trials for missing observations.

Objectives

To investigate the use of multiple imputation for a “Hypothetical” estimand and several “Treatment Policy” estimands in a NI study with an experimental and active comparator arms but no reference arm. Specifically, in the case of a treatment policy estimand the typical approach of imputing under the null hypothesis of non-inferiority [1] is compared to so-called interior family constraints [2].

Methods

Simulation study

A study design was envisaged that would asses the non-inferiority of an experimental treatment versus an active control. The sample size was determined for comparing the mean change from baseline in Hemoglobin (Hgb) concentration between the two arms. A total sample size of 2084 subjects with a 1:1 allocation ratio would provide 90% power with a two sided alpha of 0.025 using a non-inferiority margin of 0.25 g/dL.

Twenty four scenarios were considered generating missing data under a Missing at Random (MAR) and Missing Not at Random (MNAR) assumption with different discontinuation patterns and rates. Hemoglobin concentration values were generated for each subject at each visit using a multivariate normal distribution. Seven nominal study visits were generated including baseline and 6 longitudinal visits (corresponding, for example, to: 4, 8, 12, 16, 20 and 24weeks). The associated mean vector and covariance matrix used were estimated from a previously conducted phase IIB study.

Treatment discontinuation implies study discontinuation – no observations are therefore made after a treatment discontinuation. Study discontinuation for each visit assumes lack of efficacy and does so under either MAR or MNAR:

- Under MAR: probability of discontinuing at visit k+1 depends on the Hgb value at visit k.
- Under MNAR: probability of discontinuing at visit k+1 depends on the Hgb value at visit k+1

Under each scenario 10,000 simulated datasets were generated. The Hgb profiles for each dropout pattern is illustrated in Figure 1.

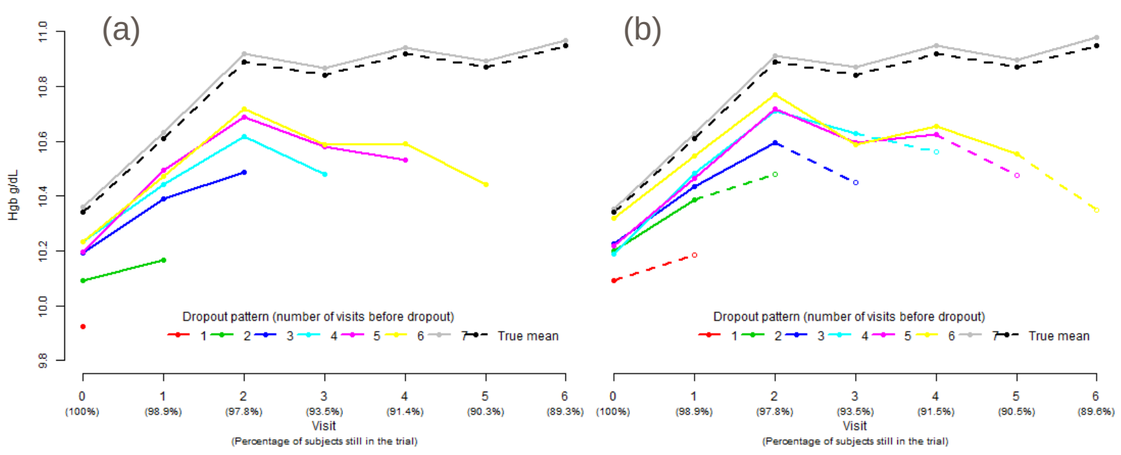

Figure 1 – Average Hgb values by dropout pattern (a) under MAR (b) under MNAR

Multiple Imputation (MI) strategies

Different assumptions were made for treatment discontinuations in each of the two arms.

Under the control arm subjects discontinuations were assumed to be MAR and imputed assuming they would have behaved like subjects in the same arm had they not discontinued.

Under the treatment arm subjects discontinuations were imputed using one of the following MI methods:

Method	Description	Strategy
Under the Null	Subjects were assumed to worsen from “MAR” by an amount of non-inferiority margin post-discontinuation.	Treatment policy
Average Case Missing Value (ACMV)	For a subject discontinuing at visit $d < 7$ the value at visit $d+1$ was imputed using information from all subjects who have completed at least $d+1$ visits. Weights $w_{d+1} > w_{d+2} > \cdots > w_7$ (adding up to 1) were given to these patterns to give more weight to the nearest patterns when imputing the value at visit $d+1$.	Treatment policy
Completer Case Missing Value (CCMV)	Imputed values were generated based on the completers only.	Treatment policy
Nearest Case Missing Value (NCMV)	Subjects with pattern discontinuation $d < 7$ had the value corresponding to visit $d+1$ imputed using the information provided by subjects in pattern $d+1$. The observed values in the first d visits and the imputed value for visit $d+1$ were used to impute the value for the visit $d+2$ borrowing information from subjects in the pattern $d+2$, and so on.	Treatment policy
Missing at Random (MAR)	Subjects were assumed to be missing at random and their missing values imputed accordingly.	Hypothetical

For each scenario, the outlined methods of analysis were performed and the operating characteristics derived including power and type 1 error.

Results

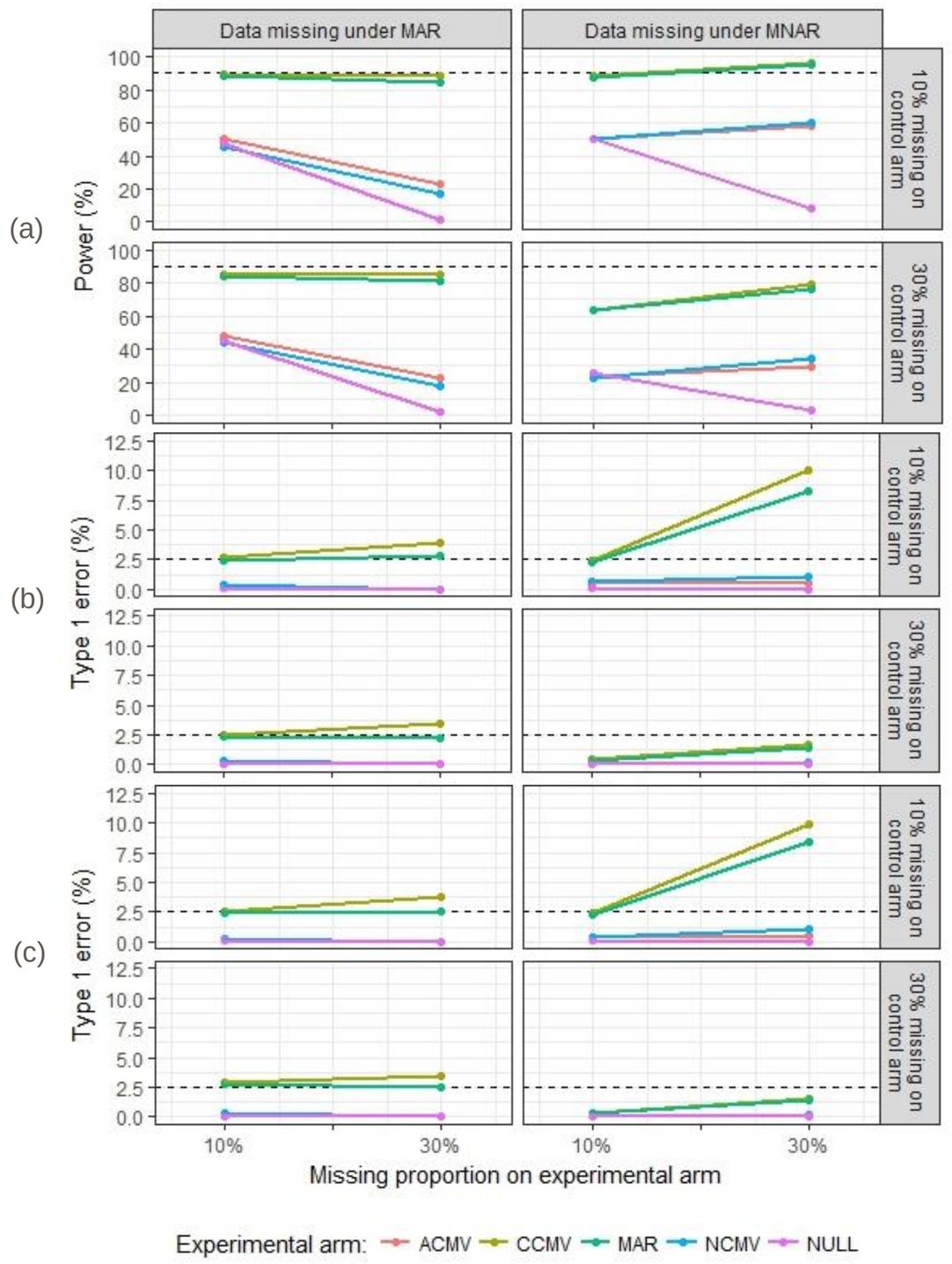

Figure 2 – Operating characteristics: (a) Power – data generated under the Alternative assumption true treatment difference = 0; (b) Type I error – data generated under the Null true treatment difference assumed to be 0 Weeks 4 & 8; NI Margin thereafter; & (c) Type I error – data generated under the Null true treatment difference gradually increased at each visit until last visit reached NI Margin

Conclusions

The “impute under the null” method was the most conservative in maintaining the Type 1 error under the nominal 2.5% threshold, though at the cost of some large losses in power.

CCMV and MAR performed similarly and fairly well with regards to power relative to the other methods. MAR kept Type 1 error at the nominal level, except in the scenario when the amount of missing data on the experimental arm was large (i.e. 30%) and on the control arm was small (i.e. 10%). CCMV performed slightly worse in preserving the type I error, especially when the amount of missing data on the experimental arm was large (i.e. 30%)

The ACMV and NCMV imputation methods performed similarly across the scenarios explored under this simulation study, and didn’t perform very well in terms of power. The NCMV method could not be applied to some simulated datasets due to insufficient numbers of subjects in each discontinuation pattern casting doubt over how feasible this method may be in practice. Under some scenarios this happened as often as in 53% of the simulated datasets.

Performance of most approaches seems more impacted by balanced vs. unbalanced discontinuation rates than by MAR vs. MNAR mechanisms of discontinuation.

Whilst ACMV, NCMV, and “impute under the null” maintained the type 1 error well beneath the nominal level, they do not seem to be very appropriate for ensuring a well powered NI study, if the missing data is due to lack of efficacy as in the simulations.

References

[1] Koch GG. (2008). Comments on ‘Current issues in non-inferiority trials’. Statistics in Medicine, 27:333–342.
[2] Kenward MG, Molenberghs G, Thijs H. (2003). Pattern-mixture models with proper time dependence. Biometrika, 90:53-71.

Acknowledgements

We would like to thank Tim Montague, Nicky Best & the GSK Missing Data Working Group for their input on this work.