Missing Data – How Much Is Too Much

Lilianne Kim, PhD[1], Kim Hung Lo, PhD[1]

[1]Janssen Research and Development, Inc., 1400 McKean Road, Springhouse, PA 19477

Abstract

For the past few years, missing data in controlled clinical trials have been a forefront issue for statisticians due to the concern of biased efficacy results from regulatory agencies. Multiple methods for analyses and data imputation have been suggested to mitigate biased results. These include multiple imputation methodology, model-based methods for example MMRM, and tipping point analyses as sensitivity analyses. But how much missing data is too much? In this proceeding, we will explore missing data for Phase III controlled clinical trials through simulations. Specifically, various degrees of missing data will be simulated to examine how much will negate a positive efficacy outcome and how much will create a biased false positive. Dichotomous endpoints were considered, and continuous endpoints may be address at a later time.

The results from simulations showed that for clinical trials with dichotomous endpoints, regardless of the types of missing data, missing data by as much as 20% of the data still retains adequate power to detect a difference between treatment groups. Furthermore, the Type I error rate was maintained for the most part.

Key Words: Missing data, dichotomous endpoints, MI

1. Background

Among the key requirements for a successful phase III confirmatory clinical study, the most relevant consideration is the statistical significance of the primary efficacy endpoint as, in most cases, the experiment is mainly designed to address the question surrounding it. Therefore, the primary efficacy endpoint needs to be carefully selected based on clinical relevancy and the ability to differentiate between active treatment and placebo groups. At the same time, actions should be taken to eliminate/minimize any 'noise' introduced during the conduct of the study. One of the major sources of 'noise' is missing data due to various reasons. Statisticians have been trying to handle this issue with various methodologies. In this article, we will focus on binary endpoints. Some of the commonly used approaches are non-responder imputation (NRI) and multiple imputations (MI).

2. Objective

The objective of this exercise was to explore how much missing data will potentially impact the success of a confirmatory clinical study. Simulations were performed on scenarios including no missing data, missing completely at random (MCAR), missing at

random (MAR), and missing not at random (MNAR) (Rubin and Little, 2002). Various analyses including MI, NRI and Complete-Case analysis were explored.

3. Methods

3.1 Simulations and Simulation Parameters

For simplicity and without loss of generality in the simulations, two treatment groups were considered, placebo and active treatment (*trt* = 0, 1, respectively). A sample size of 50 per treatment group was used with a placebo response rate, p_{pbo}, of 0.2 and an active treatment response rate, p_{trt}, of 0.5. Based on these parameters, if there is no missing data, there is about 90% power to detect a difference between the two treatment groups, with a 2-sided level of significance of 0.05.

The response, $Y = 0$ or 1, is dichotomous and is generated by using different responder proportions for the simulation scenarios. An independent variable $Z \sim N(\mu_z, \sigma_z^2)$ was generated to facilitate missing at random data and a latent variable, $X \sim N(\mu_x, \sigma_x^2)$ correlated with Y was generated for missing not at random data. The correlation between Y and X is R_{pb}.

Simulations were performed with 1000 and 5000 iterations to investigate power and Type I error, respectively.

3.2 Analyses

A logistic model was used for analyses with Z and *trt* in the model.

Data was handled 4 ways, 1) using the full sample size with no missing data, 2) Complete-Case data (no data imputation), 3) non-responder imputation (NRI), where $Y = 0$ was imputed for any missing data, and 4) multiple imputation (MI) based on a logistic model with Z and *trt* in the model.

4. Results

4.1 Power

4.1.1 Power for Missing data types

When there is no missing data, as expected, simulations confirmed that based on the parameters mentioned above, the power to detect a treatment difference is ~90%. Figures 1 a, b and c, show the reduction of the power to detect a treatment difference due to increasing proportions of missing data for the 3 types of missing data, MCAR, MAR and MNAR, respectively. For all 3 types of missing data, the Complete-Case data performed best, followed closely by MI. Power decreased at a faster rate for the NRI imputation method compared with the other two data handling methods. An interesting observation is when one looks at the top left portion of the plots for all 3 figures. Based on the plots, one can see that if there is 0% up to 20% missing data, the power to detect treatment differences is not so compromised, decreasing to ~80% except for the NRI method when data is MNAR. Beyond 20% of missing data, then the power to detect treatment differences would decrease, even by as much as 50% particularly for MNAR data.

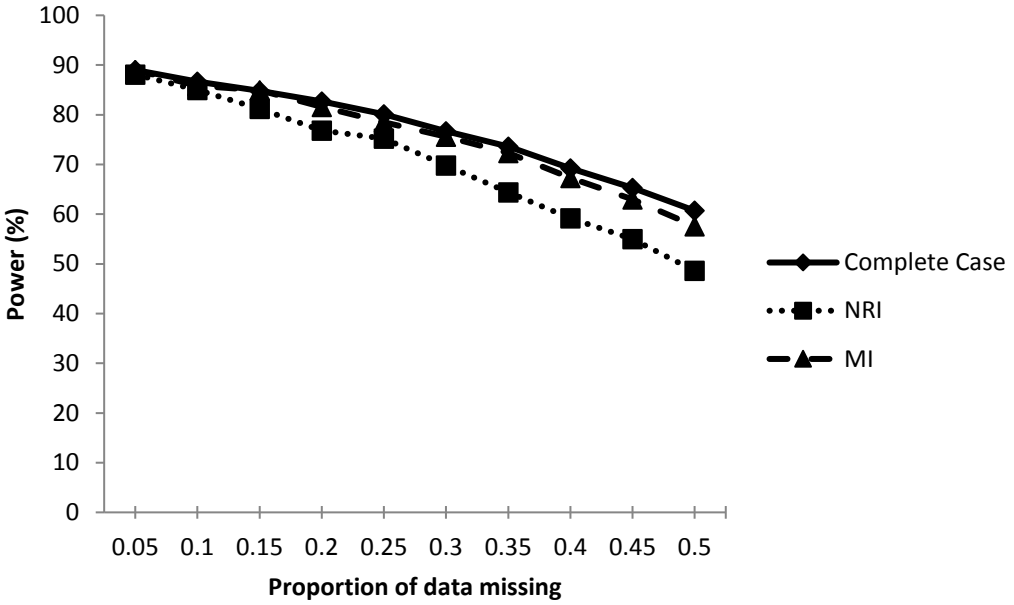

Figure 1a: Power for Missing Completely at Random data using Complete-Case, NRI and MI methods.

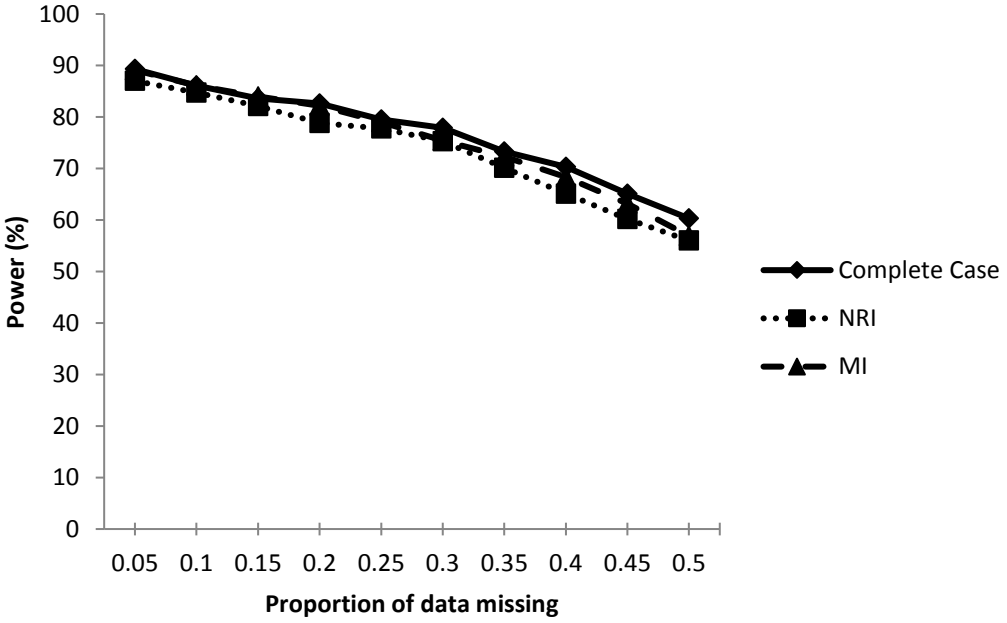

Figure 1b: Power for Missing at Random data using Complete-Case, NRI and MI methods.

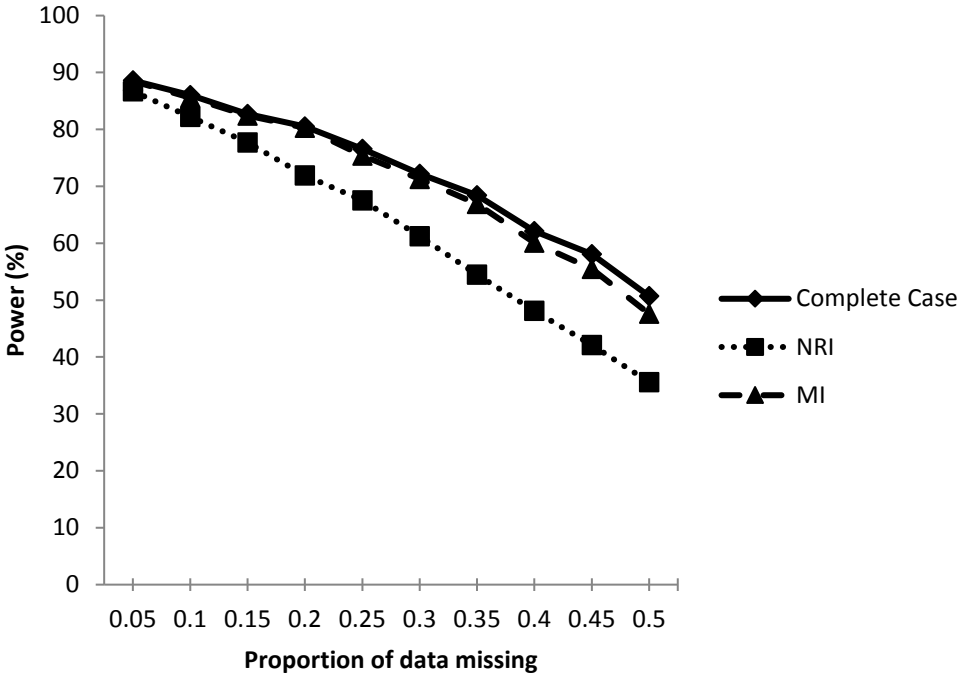

Figure 1c: Power for Missing Not at Random data using Complete-Case, NRI and MI methods.

4.1.2 Power for MNAR data for different degrees of correlation

The degree of correlation, R_{pb}, between the response, *Y,* and the latent variable, *X*, was also examined when data is MNAR. Figures 2a, b and c, shows the 3 methods of handling missing data, Complete-Case, NRI and MI, respectively, for R_{pb} = 0.2, 0.5, and 0.8. From these figures, as the correlation increases, the power can dramatically decrease, especially for the NRI imputation. For Complete-Case and MI imputation methods, the power to detect a treatment difference seems maintained above 80% when there is ~10-15% missing data.

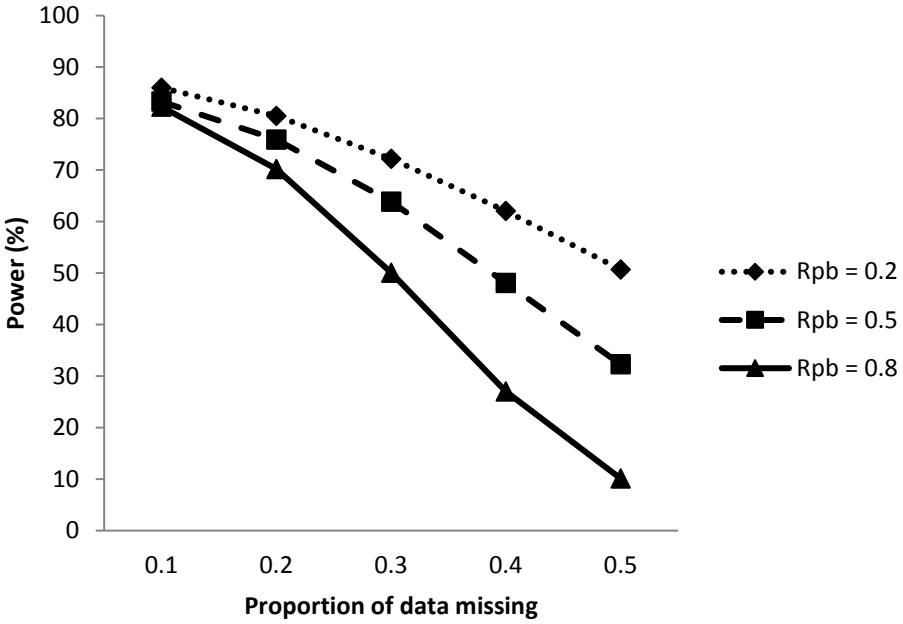

Figure 2a: Power for Missing Not at Random data using Complete-Case.

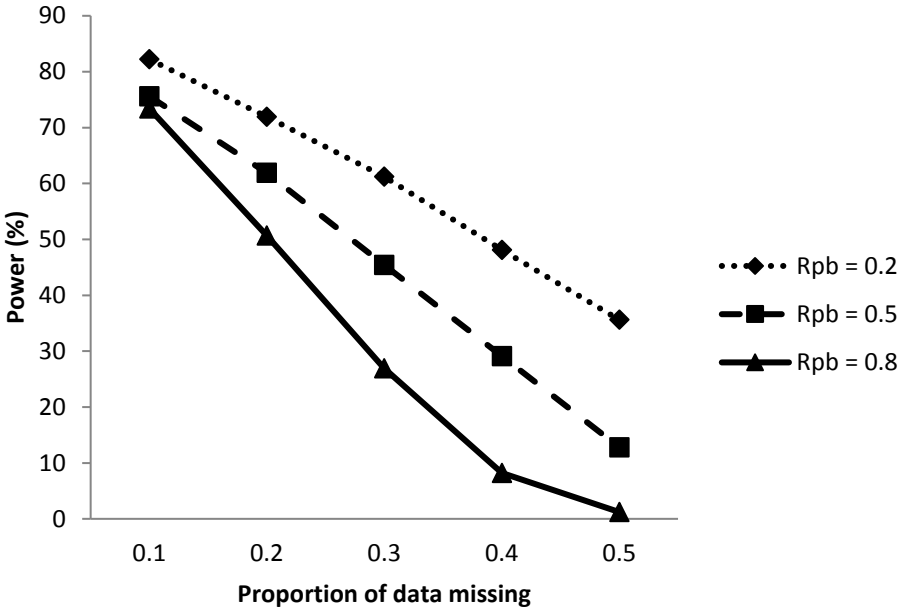

Figure 2b: Power for Missing Not at Random data using NRI.

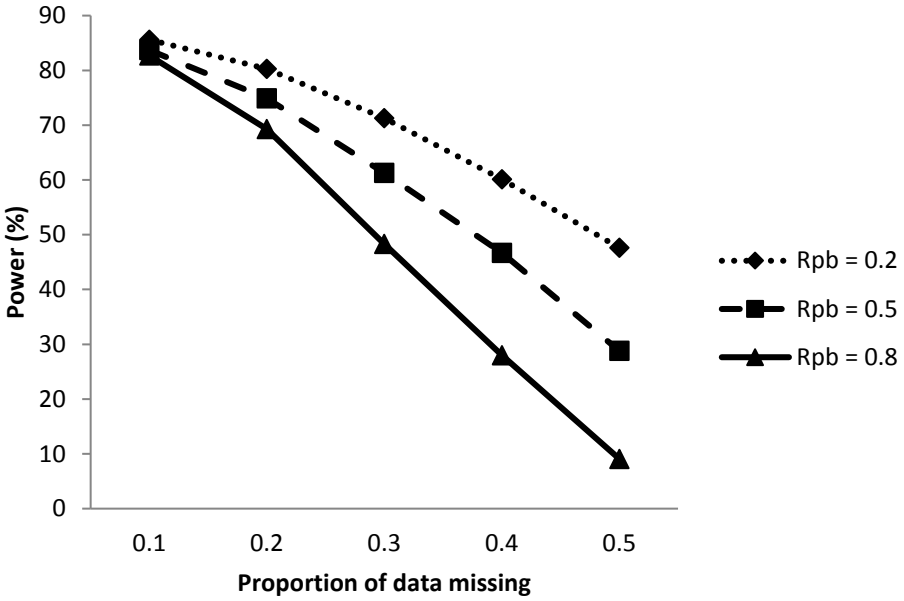

Figure 2c: Power for Missing Not at Random data using MI.

4.2 Type I Error Rate

Table 1 shows the simulation results for Type I error rates when data is MCAR, MAR and MNAR using Complete-Case, NRI and MI data handling methods. From the table, it is observed that for the most part, the Type I error rates had been retained at ~5%. The bolded scenarios represent error rates that exceed that of the scenario when there was no missing data. These were all within 10% of the no missing data case.

Table 1: Type I Error Rates for MCAR, MAR and MNAR types of missing data using Complete-Case, NRI amd MI data handling methods

Missing data type	*Pr(Y = 1)*	*Proportion missing data*	*No missing data*	*Complete-Case*	*NRI*	*MI*
MCAR	0.3	0.1	5.4	**5.52**	**5.54**	5.4
		0.2		5.3	4.98	5.2
		0.3		5.12	4.66	5.08
		0.4		4.72	4.48	4.82
		0.5		4.46	4.1	5.12
	0.5	0.1	5.2	4.68	**5.4**	4.92
		0.2		5.16	**5.26**	4.9
		0.3		4.74	5.22	5.14
		0.4		5.06	4.82	4.86
		0.5		4.74	4.82	4.98
	0.8	0.1	4.9	4.84	**5**	4.84
		0.2		4.42	4.4	4.44
		0.3		3.74	4.92	4.2
		0.4		3.4	5.36	4.04
		0.5		3.02	4.88	4.2

MAR	0.3	0.1	5.4	5.22	5.22	**5.6**
		0.2		5.28	4.94	5.3
		0.3		5.02	4.92	5.3
		0.4		4.9	4.42	4.54
		0.5		4.66	4.94	**5.6**
	0.5	0.1	5.2	4.8	4.78	4.6
		0.2		5.08	5.12	4.88
		0.3		4.74	**5.64**	4.98
		0.4		4.92	5.14	5.16
		0.5		5	5.38	5.2
	0.8	0.1	4.9	4.72	4.4	4.74
		0.2		4.48	4.58	4.58
		0.3		4.12	4.18	4.65
		0.4		3.82	4.3	4.08
		0.5		3.54	4.28	4.1
MNAR	0.3	0.1	5.4	4.82	4.48	4.74
		0.2		4.62	4.26	4.62
		0.3		3.86	3.8	4.28
		0.4		2.98	2.54	3.36
		0.5		1.74	1.1	3.36
	0.5	0.1	5.2	5.02	**5.3**	5.1
		0.2		4.92	5.16	5.12
		0.3		5.08	4.8	4.82
		0.4		4.68	4.64	5.18
		0.5		5	4	4.42
	0.8	0.1	4.9	4.58	4.88	4.7
		0.2		4.3	4.84	4.3
		0.3		4.6	**5.04**	4.8
		0.4		4.38	**5.22**	4.68
		0.5		4.48	4.56	4.58

5. Conclusions

Results from the simulations suggest that for clinical trials with dichotomous endpoints, regardless the types of missingness, missing data by as much as 20% of the data still retains adequate power to detect a difference between treatment groups. However, in the case when data is MNAR and the correlation between the latent data and the response data is high, the power to detect differences can still be maintained if there is up to ~10-15% missing data. Furthermore, the Type I error rate was maintained across the different scenarios. As a note, the best practice is always for all clinical trials to minimize missing data and 'noise' during the conduct of the study.

References

Little, RJA, Rubin DB. Statistical Analysis with Missing Data; Wiley, Sep 9, 2002.