Missing Data Handling for Phase III Trial Data: A Comparison of Single Imputation Methods

Ruishan Lin[1] Zelin Wang[1] Xiangyi Xi [2]

George Mason University[1] University of Connecticut[2]

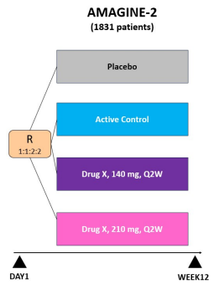
1. Overview

Background: AMAGINE-2 is a 12-week (6 visits for each patient), four-arm, randomised, double-blinded, placebo-controlled, multiple-dose phase III trial to evaluate the efficacy of drug X in treating psoriasis. In total, there are 1831 patients with mid-to-severe psoriasis randomised into four arms (placebo, active control, 140 mg of drug X every two weeks, and 210 mg of drug X every two weeks) on a 1:1:2:2 ratio.

Primary Endpoint: The primary endpoint of the study is the proportion of patients achieving 75% improvement in Psoriasis Area and Severity Index (PASI 75) at week 12.

Analytical Approach: Based on the complete dataset which is simulated from the summary statistics of the true data, we implement different missing data scenarios and analyze the PASI 75 data with logistic regression on sex and treatment.

Goal: Investigate the impact of missing data mechanisms and different levels of missingness on the result of logistic regression.

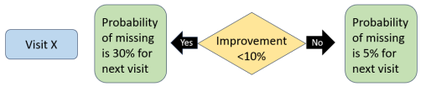


Summary of AMAGINE-2

2. Missing Data Mechanisms

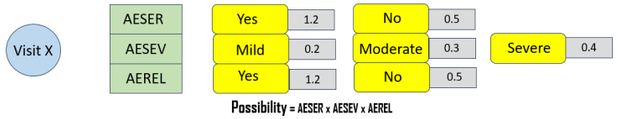
The missing data mechanisms of interest are defined as below

- **Missing Completely at Random (MCAR):** The probability of missing is the same for all patients. We explore the missing of 10%, 20%, and 30%.
- **Missing at Random (MAR):** The probability of missing at a particular visit depends on the improvement of PASI75 from baseline. Details can be found in the graph below.

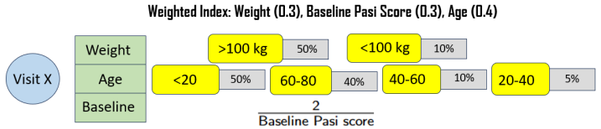


- **Missing Not at Random (MNAR):** The probability of Missing depends on unobserved predictors. We investigate the following four latent predictors:

1. A subject-level toxicity Index based on the seriousness, severity, and relatedness of adverse events.



2. A probability of missing that depends on the improvement of PASI 75 on visit 6 only
3. A probability of missing that depends on the improvement of PASI 75 on each visit.
4. A weighted probability of missing based on age, weight, and baseline PASI 75 score.



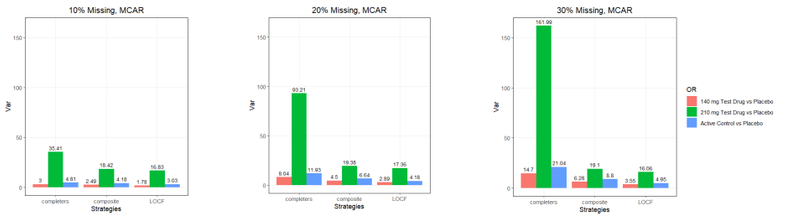
3. Approaches for Handling Missing Data

We investigate three approaches to handling missing data:

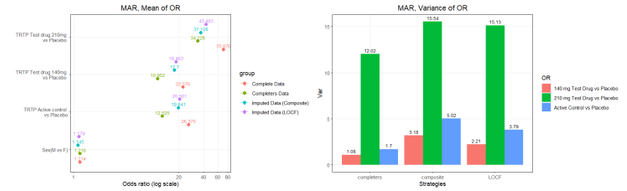
1. **Complete Case Analysis:** Analysis is performed soely on completers. Subjects with any missing data are excluded from analysis.
2. **Single Imputation with Composite Strategy:** Impute the patients who have missing data at visit 6 to be non-responders (PASI 75 = 0).
3. **Single Imputation with Last Observation Carries Forward (LOCF):** Missing data for a subject at visit 6 is replaced with the last observed non-missing value for that subject.

4. MNAR and MAR Large Sample Results

After repeating the simulation for the MNAR and MAR mechanism 1000 times, we get the asymptotic mean and variance of odds ratios for different approaches.



We can conclude that the analysis with completers approach yields the least bias from the true odds ratios. Imputation with composite strategy and LOCF both underestimate the treatment effect size, which make sense since they are conservative methods. Meanwhile, the completers approach also display the largest variation among all three methods. As missing level increases, the bias of mean odds ratio increases, so does the variance.



For the MAR scheme, we can see that LOCF approach is the least biased for the mean of odds ratio, while the completers approach deviates the most from the true results. As for variances of odds ratios, both LOCF and composite strategies yield larger variances than the completers approach.

5. MNAR Simulation Studies

Predictor	group vs placebo	Mean of Odds Ratios			
		Complete	Completer	Composite	LOCF
Toxicity Index	Active Control	26.38	25.51	18.62	19.76
	Test Drug 140mg	22.58	22.16	18.06	18.94
	Test Drug 210mg	70.88	72.85	37.09	44.77
Improvement in PASI	Active Control	26.38	22.09	19.62	19.46
	Test Drug 140mg	22.58	19.73	18.40	18.15
	Test Drug 210mg	70.88	64.10	46.36	45.74
Cumulative Improvement in PASI	Active Control	26.38	25.78	18.71	19.01
	Test Drug 140mg	22.58	22.28	16.84	17.45
	Test Drug 210mg	70.88	70.41	34.02	39.56
Weight, Age and Baseline	Active Control	26.38	24.85	12.47	13.05
	Test Drug 140mg	22.58	23.91	12.05	13.02
	Test Drug 210mg	70.88	73.49	18.21	24.44

MNAR: Large Sample Mean of Odds Ratio

Predictor	group vs placebo	Variance of Odds Ratios		
		Completer	Composite	LOCF
Toxicity Index	Active Control	2.84	2.13	1.91
	Test Drug 140mg	1.65	1.4	1.14
	Test Drug 210mg	23.47	9.16	10.43
Improvement in PASI	Active Control	6.4	3.56	2.9
	Test Drug 140mg	3.75	2.2	1.66
	Test Drug 210mg	56.58	18.3	14.9
Cumulative Improvement in PASI	Active Control	10.43	5.85	4.14
	Test Drug 140mg	6.53	3.9	2.64
	Test Drug 210mg	74.67	18.23	16.85
Weight,Age, and baseline	Active Control	28.76	6.82	3.51
	Test Drug 140mg	23.93	5.8	2.99
	Test Drug 210mg	264.83	14.01	11.4

MNAR: Large Sample Variance of Odds Ratio

Including only completers yields the most accurate result for all four MNAR schemes. However, the variance of odds ratios is the highest when we include only completers.

6. Discussions

- In this project, we are only concerned with the odds ratios at visit 6. If one wishes to incorporate information during visits 2-5, Cox proportional hazard model is a better choice.
- The three approaches we investigated are the simple ones. We acknowledge that there are more suitable imputation methods such as likelihood based methods, inverse probability weighting, and multiple imputations.
- It is recommended in several references that sensitivity analysis should be conducted when exploring missing data, especially for MAR and MNAR.

7. References

[1] Dziura JD et al. (2013). Strategies for dealing with missing data in clinical trials: from design to analysis. *Yale J Biol Med*, 86(3):343-58.

[2] Little et al. (2012). The prevention and treatment of missing data in clinical trials. *The New Engliand Journal of Medicine*.

[3] Little et al. (2002). *Statistical Analysis with Missing Data, Second Edition*. Wiley.

[4] Mallinckrodt et al. (2020). *Estimands, Estimators and Sensitivity Analysis in Clinical Trials*. Routledge.

[5] Schmidli. *Causal Reasoning and Strategies for Defining Estimands*. Novartis.