Estimating Treatment Effects in Longitudinal Clinical Trials with Missing Data

A Preprint

Amy Browne
Department of Mathametics
University of Galway

k.yip2@universityofgalway.ie

Tsz Mang Yip
Department of Mathametics
University of Galway

a.browne47@universityofgalway.ie

June 2, 2025

Abstract

Enter the text of your abstract here.

***K*eywords** Missing Data · Longitudinal Data · Randomized Clinical Trial · Real World Data Analysis

1 Introduction

2 Project Aims

3 Background

Missing data in clinical research

3.1 Missing Data Mechanisms

Missing data mechanisms are important to consider when choosing which sort of missing data handling method to use. There are three mechanisms which missing data can follow:

- Missing Completely At Random (MCAR)
- Missing At Random (MAR)
- Missing Not At Random (MNAR)

Although they may appear similar at first glance, handling missing data without considering these mechanisms may result in biased estimates and inaccurate conclusions.

Missing Completely At Random

The formal definition of MCAR data is:

$$P(R = 1 \mid Y, X) = P(R = 1)$$

where R is a missing data indicator (1 = Data is observed, 0 = Data is missing), Y represents the variables in which the data is potentially missing and X represents the observed data. The probability of the data is observed given observed data and missing data is the same as the probability of being observed without the given data. This mechanism is considered the easiest to deal with as it does not bias the result although data is rarely MCAR. This can occur due to system failure and some data is deleted accidentally, or else there is issues with the treatment system and data cannot be recorded.

Missing At Random

$$P(R = 1 \mid Y, X) = P(R = 1 \mid X)$$

The probability of data being observed given the rest of the data is the same as the probability being observed given the observed data. In short, the data's missingness is dependent on the observed data. For example, people with a higher body mass index may be more prone to having missing blood pressure data - this is not relative to the missing data. MAR is a more realistic mechanism than MCAR and requires more intensive handling methods.

Missing Not At Random

$$P(R = 1 \mid Y, X) \neq P(R = 1 \mid X)$$

The probability of data being observed is not dependent on the observed data. This mechanism is the most difficult to deal with as it relates to the unobserved data, so producing valid results is a challenge. Certain participants in a general health study may avoid answering questions truthfully about smoking habits or their diet in order to make themselves more appealing. Sensitivity analysis is an option to determine the treatment effect when assuming different mechanisms.

4 Data

4.1 Acupuncture for chronic headache in primary care: large, pragmatic, randomised trial (Vickers et al., 2004)

4.2 The Effects of Vitamin D and Marine Omega-3 Fatty Acid Supplementation on Chronic Knee Pain in Older US Adults: Results From a Randomized Trial (MacFarlane et al. 2020)

5 Methods

5.1 Complete Case Analysis

5.2 Single Imputation

5.3 Maximum Likelihood

5.4 Multiple Imputation

6 Result

7 Discussion