Estimating Treatment Effects in Longitudinal Clinical Trials with Missing Data

A Preprint

Amy Browne *
Department of Mathametics
University of Galway

a.browne47@universityofgalway.ie

Tsz Mang Yip
Department of Mathametics
University of Galway

k.yip2@universityofgalway.ie

June 13, 2025

Abstract

Motivation. Missing data is a pervasive issue in longitudinal clinical trials, risking bias and reduced power. This study compares several statistical methods for estimating treatment effects in the presence of missing data.Result. No meaningful differences between methods for both data sets. To understanding there are many ways to handle missing data, and it is crutial to understand the mechanism to make corret choise under circumstances.Supplement information. Code Available at `https://github.com/amydebrun/Missing-Data`

Keywords Missing Data · Longitudinal Data · Randomized Clinical Trial · Real World Data Analysis · Sensitivity Analysis · Multiple Imputation · Missingness Mechanisms

1 Introduction

1.1 Missing data in longitudinal study

Longitudinal Study Design

Longitudinal studies are research designs in which data are collected from the same participants at multiple time points. They provide valuable insights into how variables change over time—often over several years or decades. While this design is powerful for understanding causal relationships and long-term trends, it presents significant challenges in retaining participants across all waves of data collection.

Longitudinal trials require consistent data collection methods at predefined intervals. Researchers must work to minimize participant dropout or non-response. However, due to the extended duration and complexity of these studies, missing data is almost inevitable.

Similarly, clinical trials which evaluate the effects of biomedical or behavioral interventions are also vulnerable to missing data. Participants may miss follow-ups due to illness or other commitments. This loss of data can compromise the study's validity. Ignoring missing data introduces bias, as we're disregarding potentially informative cases. For instance, if participants drop out due to adverse effects from the treatment, the missing data is not random—this introduces attrition bias, which can distort estimates of the treatment's effect.

Removing incomplete cases reduces the sample size, which in turn decreases statistical power, limiting the ability to detect valid treatment effects.

*The authors would like to thank Neil O'Leary from the University of Galway for his valuable guidance, feedback, and support throughout the development of this project

In statistical software like R, missing values are represented as NA. When fitting models like linear regression, R automatically excludes any subjects with missing values in either predictor or outcome variables. While convenient, this method can lead to selection bias if the missingness is systematic.

1.2 Project Objective

The objective of this project is to estimate treatment effects using real-world data, with a primary focus on addressing issues that come with missing data. We will examine different assumptions about missing data mechanisms—such as Missing Completely at Random, Missing at Random, and Missing Not at Random and investigate their implications for statistical analysis.

While a range of methods are available to handle missing data, they all vary in levels of complexity and assumptions. This project aims to explore these methods, understand the reasons behind their differing performance, and implement them on real-world longitudinal clinical data to compare their effectiveness in estimating treatment effects.

After initial exploration of the different methods, we also decided to adjust the estimands through data wrangling our chosen datasets to compare how the estimate differs when the estimand is either categorical or continuous and to compare the effects using linear regression or linear mixed effects models. Additionally, we will conduct sensitivity analyses to investigate how treatment effect estimates change under different assumptions about the missing data.

1.3 Missingness Mechanisms - unverifiable, mcar little test

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, continuing to handle missing data without considering these mechanisms may still result in biased estimates and inaccurate conclusions. Missing data mechanisms appear to be mentioned first by Donald Rubin in 1976. We can formally define these mechanisms using the following notation;

- R represents a missing data indicator which when $R = 1$ indicates observed data and $R = 0$ indicates unobserved
- X represents data that is always observed
- Y represents data that is potentially missing.

Missing Completely At Random

The formal definition of MCAR data is:

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

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.

-insert dag

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.

-insert dag

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.

- DAG of missing mechanism (randomised) in progress in R scripts

1.4 Different missing data patterns (section tbc)

A missing data pattern describes the pattern of missing data among the observed data. As described by Little & Rubin (2014), it's important to remark the missing data patterns of the data set prior to data handling as some handling methods are intended for certain classified patterns.

```
## # A tibble: 6 x 2
##   `Missing Data Pattern` Description
##   <chr>                  <chr>
## 1 Univariate             "Missing values are confined to a single variable"
## 2 Multivariate           "Missing values can be present in multiple variables"
## 3 Monotonic              "variables $Y_j$ can be ordered if missing, with varia~
## 4 General                "Missing values appears to have no structure and scatt~
## 5 File Matching          "Missing data appears to be statistically matched"
## 6 Factor Analysis        ""
```

1.5 Literature and Limitations

Current literature

(Spineli et al. 2015) - after 2009 - 190 systematic reviews - 175 contained missing data in one study - Randomised controlled trial - 35% reported implications of missing data - 78% performed intention to treat/ including trials imputed outcome data - less than 20% performed sensitivity analysis

(Hunt et al 2021) - systematic review 2018 and 2019 - 62 studies; 2726 articles - 56% studies reported missing data - 11 reported potntial bias - 19 reported missing data handling methid - 13 Complete case - 2 multiple imputation - 2 used CCA and MI - 1 Mean imputation - 1 imputed similar variable info

- pharmacoepidemiologic multi-database studies

(Baker et al. 2025) - 229 studies observational studies UNOS - 78 reported how they use missing data - CCA common (41) - multiple imp (22) - other methods (15) - 31 removed covariates from analysis

Existing literature reveals inconsistent reporting and handling practices.(Power 2014)

2 Real World Data

We have sourced two different data sets to perform our missingness analysis on. They both have different sample sizes and different levels of missing data which is an advantage as we can conduct analysis in different settings.

Acupuncture Data

Source

Vickers et al. published a paper in 2004 on a trial they conducted to determine the effect of acupuncture therapy on chronic headache in primary care. Vickers released the dataset from the study in 2006 and is publicly available to download in excel format from `https://pmc.ncbi.nlm.nih.gov/articles/PMC1489946/`.

Study Design

The study design for the acupuncture trial is a longitudinal randomised controlled trial with two follow up time points. 401 participants were gathered from general practices in England and Wales who suffered from migraines.

Data

13 variables were recorded during the acupuncture trial.

```
## # A tibble: 14 x 2
##    Variable      Description
##    <chr>         <chr>
##  1 id            patient ID code
##  2 age           Age
##  3 sex           sex; female (1) vs. male (0)
##  4 migraine      diagnosis ; migraine (1) vs. tension-type (0)
##  5 chronicity    number of years of headache disorder
##  6 acupuncturist acupuncturist id code
##  7 practice_id   gp practice id
##  8 group         treatment group; acupuncture (1) vs. control (0)
##  9 pk1           headache severity score baseline
## 10 pk2           headache severity score 3 month
## 11 pk5           headache severity score 1 year
## 12 f1            headache frequency baseline
## 13 f2            headache frequency 3 month
## 14 f5            headache frequency 1 year
```

Variables `pk2` and ‘`pk5` are the follow-up times at 3 months and 12 months.

```
##
##
## |**Characteristic** | **0**  N = 196 | **1**  N = 205 |
## |:------------------|:--------------:|:--------------:|
## |id                 |   470 (209)    |   472 (206)    |
## |age                |    45 (11)     |    46 (11)     |
## |sex                |    165(84%)    |    172(84%)    |
## |migraine           |    183(93%)    |    194(95%)    |
## |chronicity         |    22 (13)     |    21 (14)     |
## |acupuncturist      |  5.97 (2.82)   |  6.00 (2.80)   |
## |practice_id        |    24 (11)     |    25 (12)     |
## |pk1                |    27 (17)     |    26 (15)     |
## |pk2                |    24 (18)     |    19 (16)     |
## |Missing            |       43       |       32       |
## |pk5                |    22 (17)     |    16 (14)     |
## |Missing            |       56       |       44       |
## |f1                 |     16 (7)     |     16 (7)     |
## |f2                 |     11 (9)     |     11 (8)     |
## |f5                 |     10 (9)     |     9 (8)      |
```

Figure 2.x is the summary statistics of the acupuncture trial data. Note pk2 and pk5 are post randomisation variables and contain missing data.

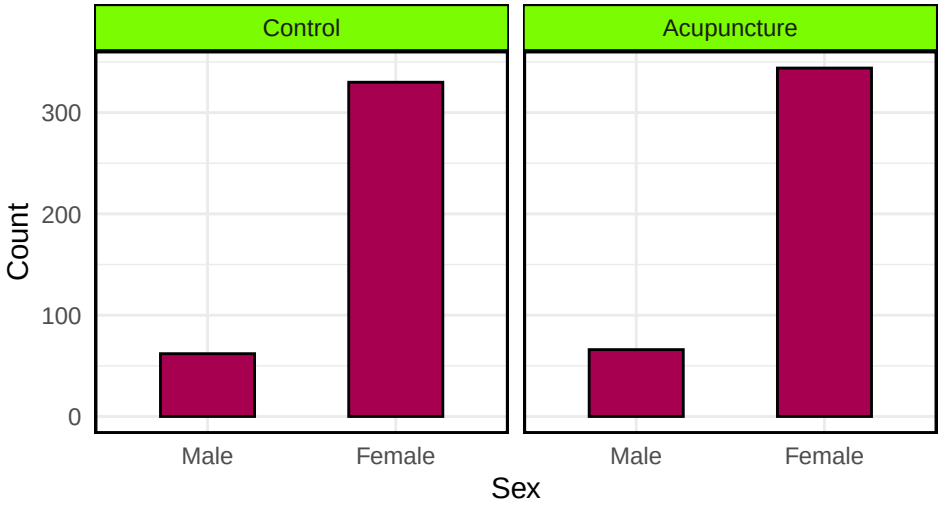


Figure 2.x is the distribution of sex in both treatment groups of the acupuncture trial data. The participants are predominantly female.

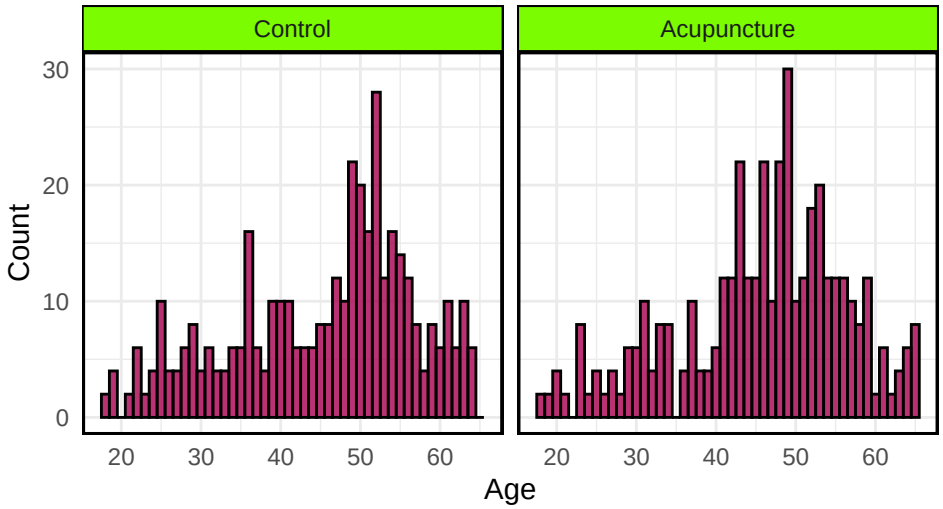
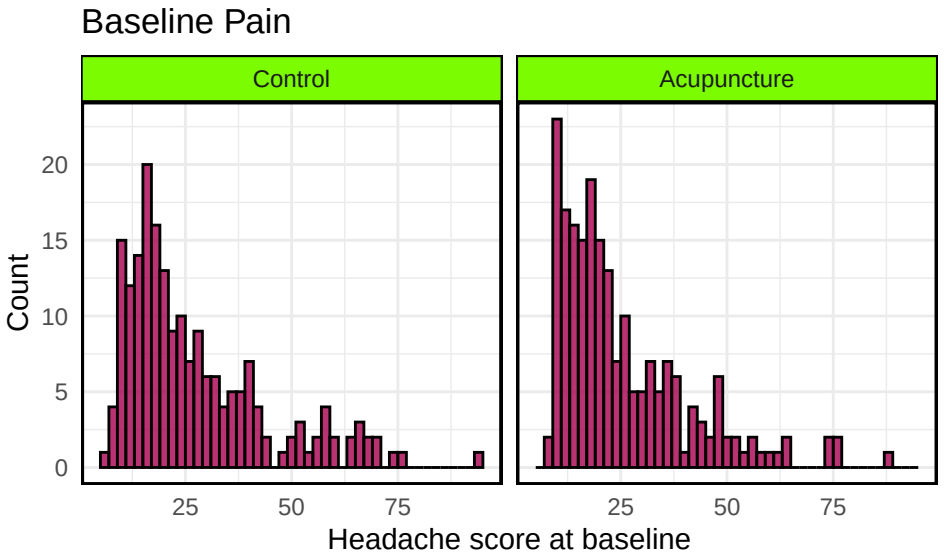
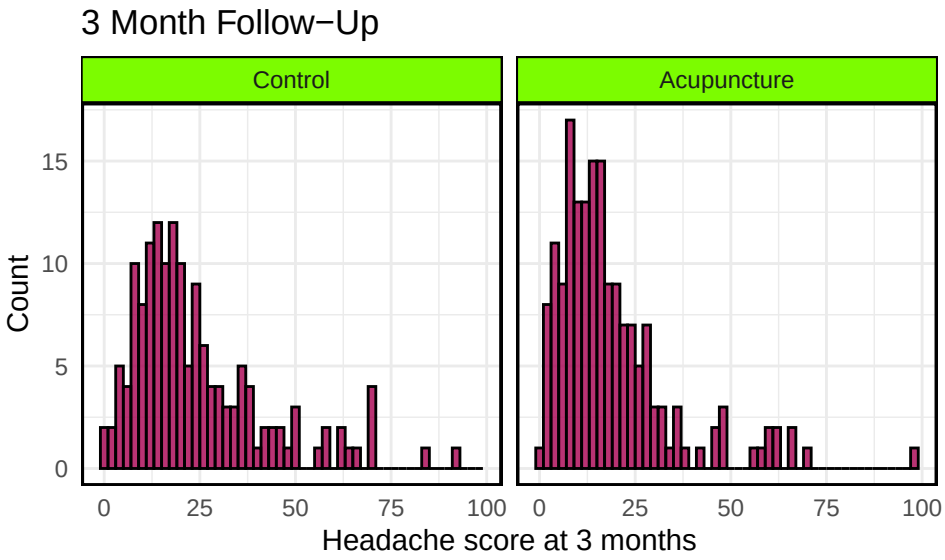


Figure 2.x is the distribution of age in both treatment groups. Both distributions are negatively skewed, predominantly middle aged and older.


Baseline Pain
Control
Acupuncture
Count
0
5
10
15
20
25
50
75
Headache score at baseline


3 Month Follow−Up
Control
Acupuncture
Count
0
5
10
15
0
25
50
75
100
Headache score at 3 months

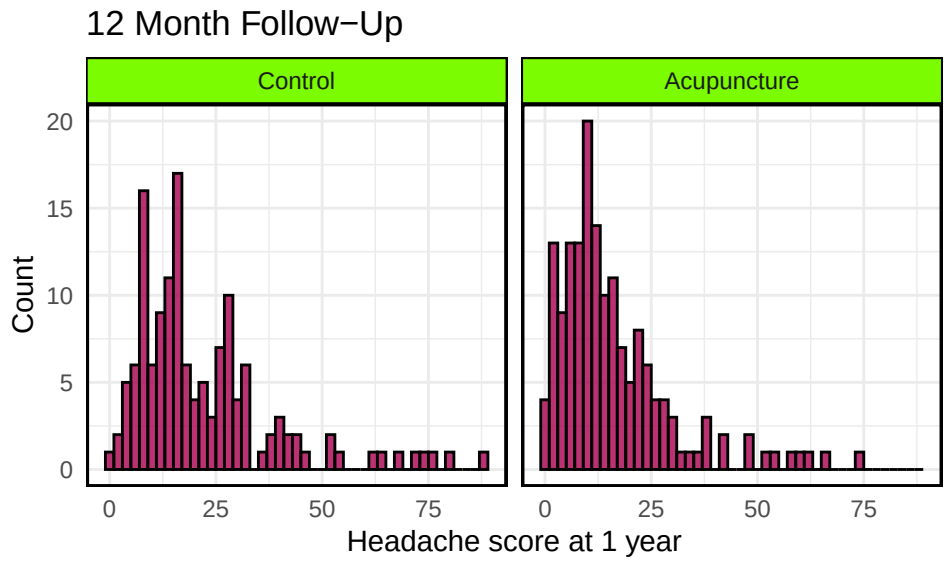


Figure 2.x,2.x,2.x: These plots show the distribution of the headache severity score at the three measured time-points. Over the three time-points we can see a generally positively skewed distribution with most headache severity scores ranging between 0 and 35. There appears to be minimal difference in headache severity score over the treatment group, and the control group appears to have a lower headache severity score overall. The 3 month and 12 month time-point distributions are incomplete due to the missing values.

Original Estimand

The aim of the original study was to estimate the effect of acupuncture therapy on chronic headache in comparison to general care. Therefore, our estimand with the sourced data set is the difference in mean headache score at the final time point of 12 months between the acupuncture group and the control group. This is modelled parsimoniously by

$$\text{pk5}_i = \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{pk1}_i + \varepsilon_i$$

where

- $pk5_i$ is the headache severity of individual i at 12 months.
- β_0 is our intercept parameter when there is no treatment group and no recorded baseline headache severity.
- β_1 is the effect of the treatment group. This is our treatment effect of interest.
- β_2 is the effect of the baseline headache severity
- ε_i is the residual error term for individual i

VITAL data - missing baseline data (vital)

Source

MacFarlane et al. conducted a study on the therapeutic effects of vitamin D and fish oil on osteoarthritic knee pain. We sourced this data at xyz.

Study Design

This study was a 2x2 factorial designed longitudinal randomised controlled trial. 1,398 participants in the study was randomised to 4 possible treatment combinations. Participants were randomised to receive either omega-3 oils or vitamin D, both treatments, or randomised to the placebo group. For this research project, we assumed that that the participants were only randomised to either receive the fish oil treatment, vitamin D treatment or the placebo. We did not include the interaction between vitamin D and fish oil in our substantive model or our imputation model.

	Vitamin D	Placebo
Fish Oil	N = 342 (not analysed)	N = 353
Placebo	N = 332	N = 371

Overview

There are x variables recorded in the VITAL data, the following table refers to a subset of the variables that we frequently use.

```
## # A tibble: 13 x 2
##    Variable       Description
##    <chr>          <chr>
##  1 Subject_ID     Patient ID code
##  2 age            Age of patient
##  3 bmi            Body mass index of patient
##  4 sex            Sex of patient
##  5 vitdactive     1=vitamin D, 0=no vitamin D
##  6 fishoilactive  1= fish oil, 0= no fish oil
##  7 pain_base      Knee pain at baseline
##  8 pain_yrX       Knee pain X years post randomisation
##  9 stiffness_base Knee stiffness at baseline
## 10 stiffness_yrX  Knee stiffness X years post randomisation
## 11 function_base  Knee function at baseline
## 12 function_yrX   Knee function X years post randomisation
## 13 kneepainfreq   Frequency of knee pain
```

Characteristic	**Both** N = 342
Sex 1-male,2-female	
1	111(32%)
2	231(68%)
Body mass index at randomization, kg/m2	32 (7)
Missing	8
Current smoking 1-yes,0-no	26(7.8%)
Missing	7
Age at randomization to VITAL study,years	67 (7)
Baseline Aspirin use 1-yes,0-no	156(46%)
Missing	6
WOMAC Pain score at baseline	37 (19)
Missing	48
WOMAC Pain score at year 1	32 (19)
Missing	182
WOMAC Pain score at year 2	32 (20)
Missing	47
WOMAC Pain score at year 3	32 (21)
Missing	78
WOMAC Pain score at year 4	30 (20)
Missing	129
Have you had a knee replacement surgery? (1=Yes, 2=No)	
1	43(13%)
2	293(87%)
Missing	6
Unilateral knee pain 1=Yes 0=No	200(71%)
Missing	61
Bilateral knee pain 1=Yes 0=No	81(29%)
Missing	61
Frequency of knee pain 1=Never 2=<1 day/wk 3=1~2 days/wk 4=3~6 days/wk 5=daily	
1	1(0.3%)
2	16(5.4%)
3	44(15%)
4	58(20%)
5	169(57%)

```
## |9                                                                    |      7(2.4%)
## |Missing                                                              |        47
## |Tylenol use at baseline 1=never 2=occational 3=daily 9=missing       |
## |1                                                                    |     108(37%)
## |2                                                                    |      91(31%)
## |3                                                                    |      53(18%)
## |9                                                                    |      43(15%)
## |Missing                                                              |        47
## |Nsaids use at baseline 1=never 2=occational 3=daily 9=missing        |
## |1                                                                    |      62(21%)
## |2                                                                    |      98(33%)
## |3                                                                    |     110(37%)
## |9                                                                    |     25(8.5%)
## |Missing                                                              |        47
## |Stronger med use at baseline 1=never 2=occational 3=daily 9=missing  |
## |1                                                                    |     178(60%)
## |2                                                                    |      40(14%)
## |3                                                                    |      46(16%)
## |9                                                                    |      31(11%)
## |Missing                                                              |        47
## |TKR 1=yes 0=No                                                       |      64(19%)
```

Figure 2.x shows the summary statistics of the VITAL data. It provides information on all 4 treatment groups. Note here that the VITAL data contains missing baseline data as well as post randomisation data.

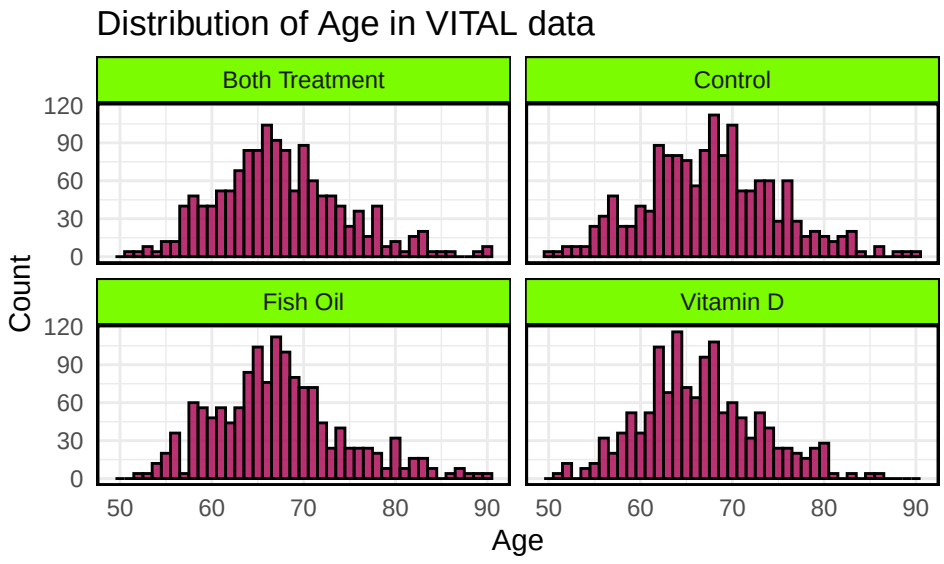


Figure 2.x shows the distribution of age in each treatment group, which appear to be normal/ positive skewed distributions.

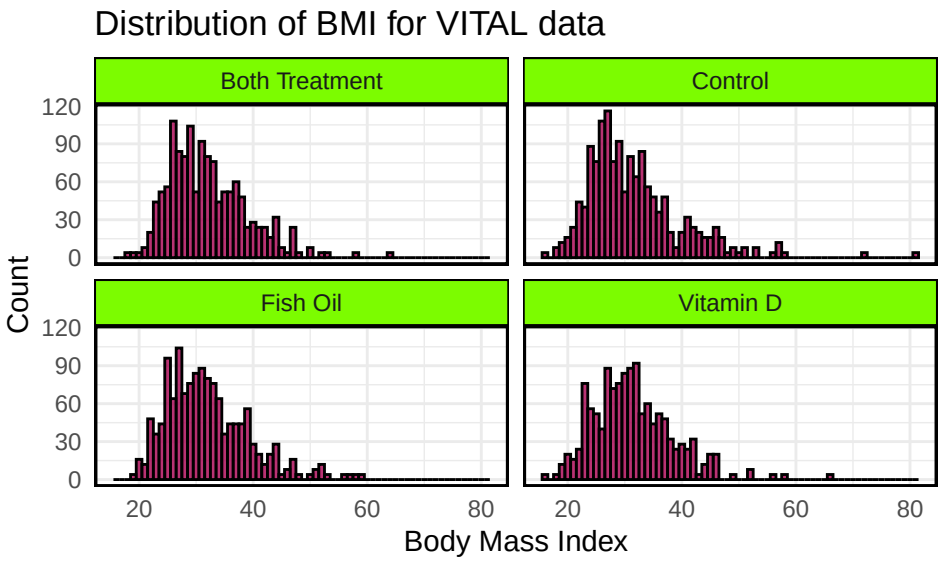


Figure 2.x shows the distribution of body mass index in each treatment group, these appear to be positively skewed with a bmi between 20 and 40. This shows a range between healthy weighted and obese participants.

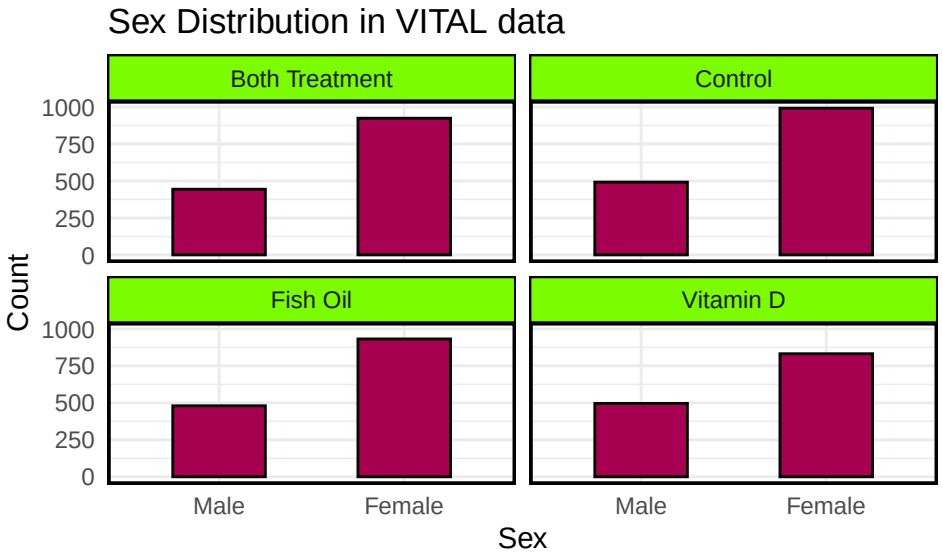
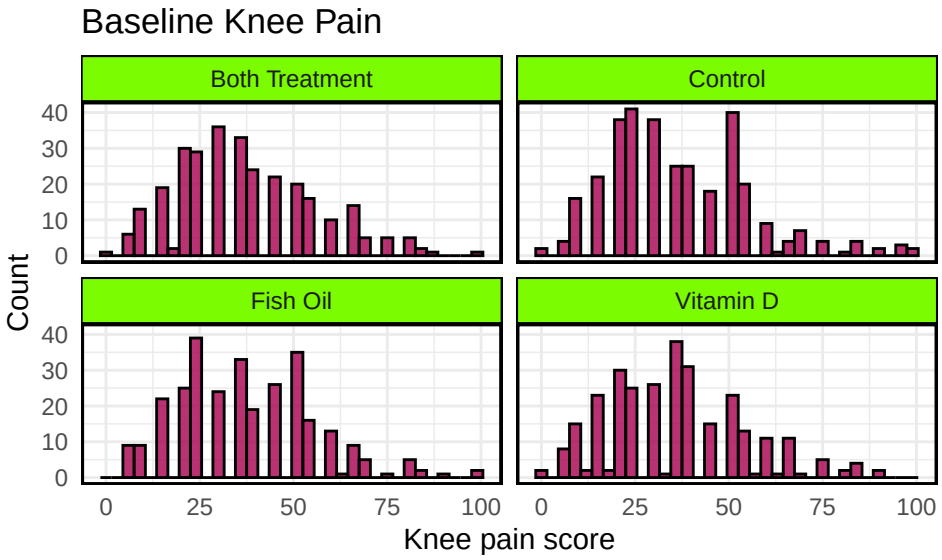
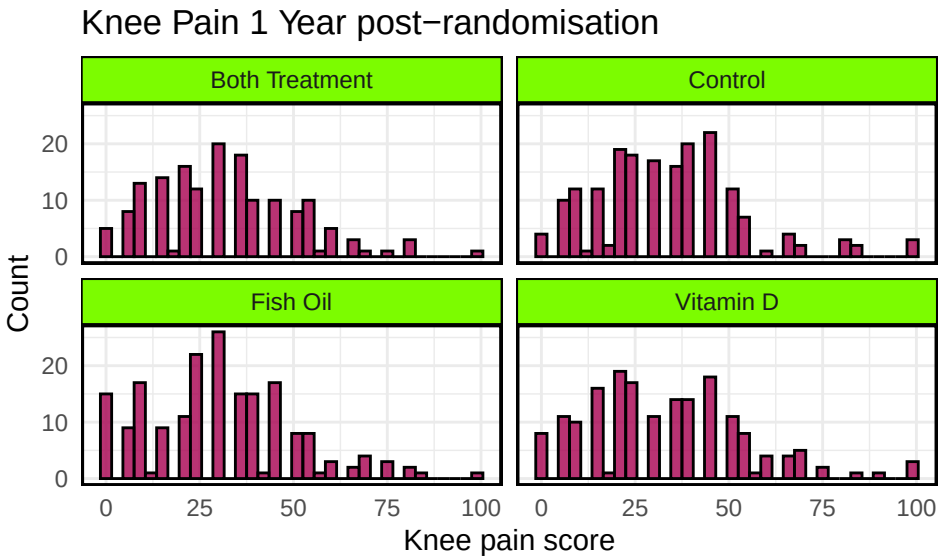
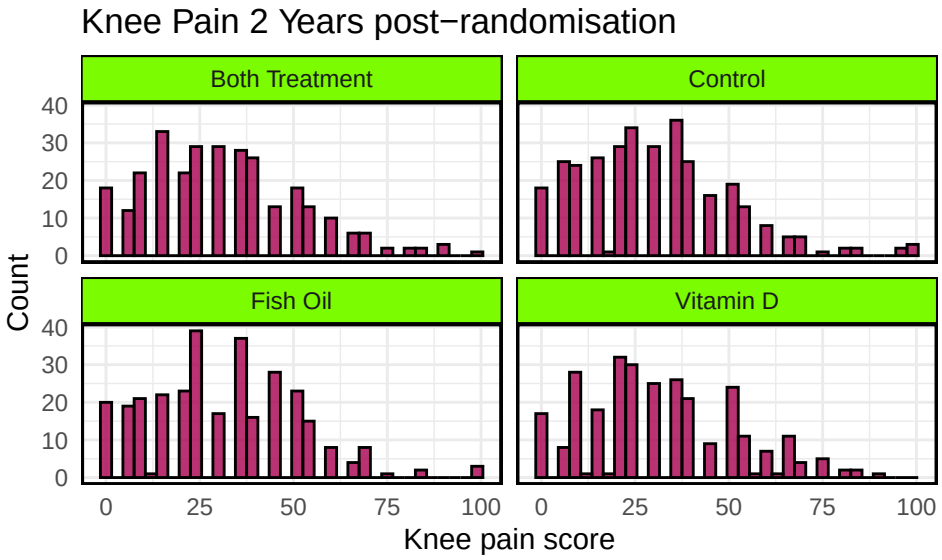
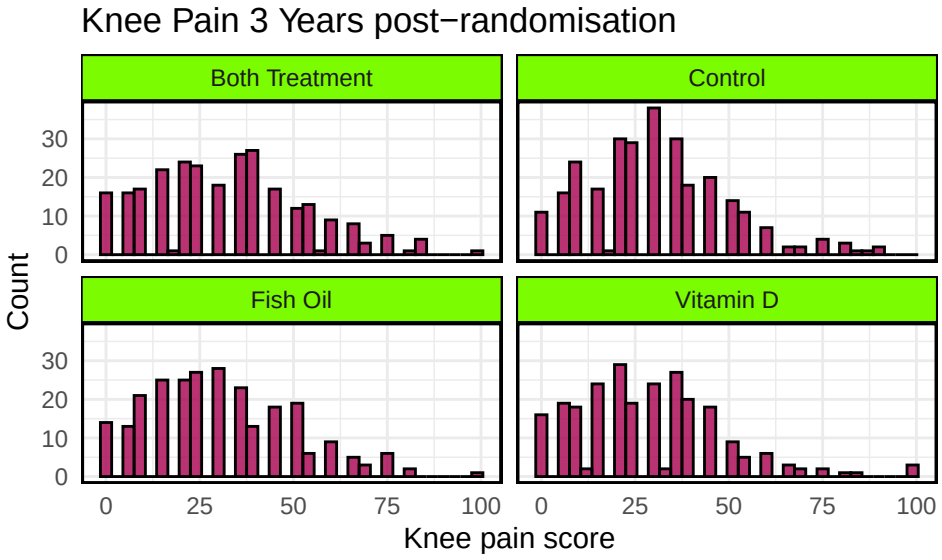


Figure 2.x is the distribution of males and females in the VITAL data. Similar to the acupuncture trial, the participants are predominantly female.


Baseline Knee Pain
Both Treatment
Control
Fish Oil
Vitamin D
Count
0
10
20
30
40
0
25
50
75
100
Knee pain score


Knee Pain 1 Year post−randomisation
Both Treatment
Control
Fish Oil
Vitamin D
Count
0
10
20
0
25
50
75
100
Knee pain score


Knee Pain 2 Years post−randomisation
Both Treatment
Control
Fish Oil
Vitamin D
Count
0
10
20
30
40
0
25
50
75
100
Knee pain score


Knee Pain 3 Years post−randomisation
Both Treatment
Control
Fish Oil
Vitamin D
Count
0
10
20
30
0
25
50
75
100
Knee pain score

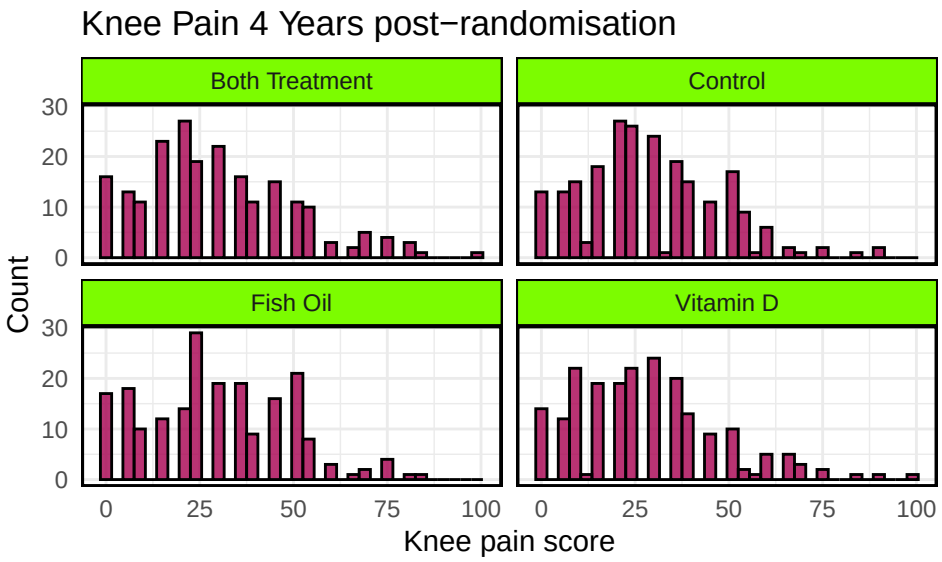


Figure 2.x,2.x,2.x,2.x,2.x shows the distribution of the knee pain from baseline until 4 years post randomisation. In all treatment groups it appears that the distributions follow a positive/ near normal distribution and skews more positive as time goes on. This occurs in all groups. Each time-point has an incomplete distribution due to missing data.

Original estimand in VITAL study

Our estimand for this dataset is the difference in mean knee pain at the final time point of 4 years post randomisation $pain_yr4$ which is modelled by

$$painyear4_i = \beta_0 + \beta_1 fishoilactive_i + \beta_2 vitdactive_i + \beta_3 painbase_i + \varepsilon_i$$

in which

- $textpain_yr4_i$ is the knee pain score for individual i at the final time point of 4 years post randomisation.
- β_0 is the intercept parameter which represents the knee pain score when there is no treatment of vitamin D or fish oil given and no baseline knee pain score is recorded.
- β_1 is the effect of being randomised to the fishoil treatment group. This is the treatment effect of interest in our fishoil only missingness analysis.
- β_2 is the effect of being randomised to the vitamin D treatment group. This is the treatment effect of interest in our vitamin D only missingness analysis.
- β_3 is the effect of the baseline knee pain score pre-randomisation.
- ε is the residual error of individual i

2.1 Missing analysis

Acupuncture Data

The summary statistics of the acupuncture trial data shows that all baseline variables have been observed. It appears two post-randomisation outcome variables contain missing data. This pattern would be described as a multivariate monotonic pattern as the percentage of missing data increases at each follow up time point, which is common in longitudinal studies. We can visualise this in the plot below.

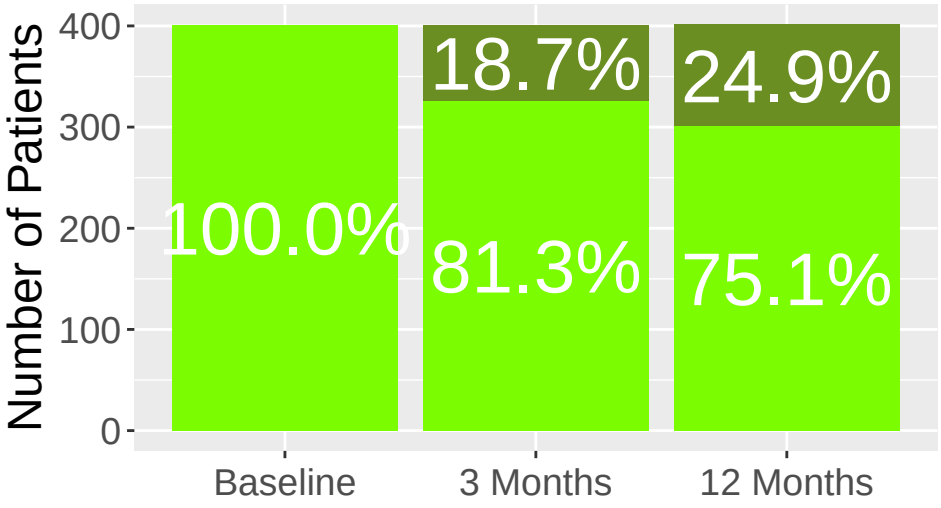


Figure 2.x: Proportion of missing data (dark green) in outcome variables $pk2$ and $pk5$. This follows a monotonic missing data pattern.

If we separate the acupuncture trial data into the treatment group and control group, we can observe that there is a higher percentage of missing data in the control group.

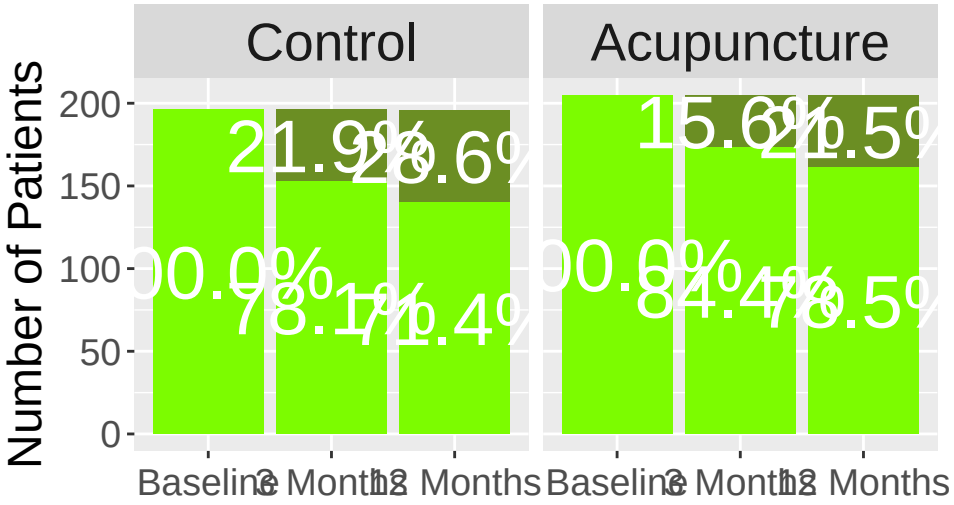


Figure 2.x: Higher percentage of missing data in control group

If we investigate the missing data pattern in more detail, we can see that the pattern is similar in both groups. In the plot below, we can see that the pattern in which the three headache severity scores $pk1$, $pk2$ and $pk5$ are observed (1,1,1) followed by the pattern in which both post-randomised headache severity scores contain missing data (1,0,0).

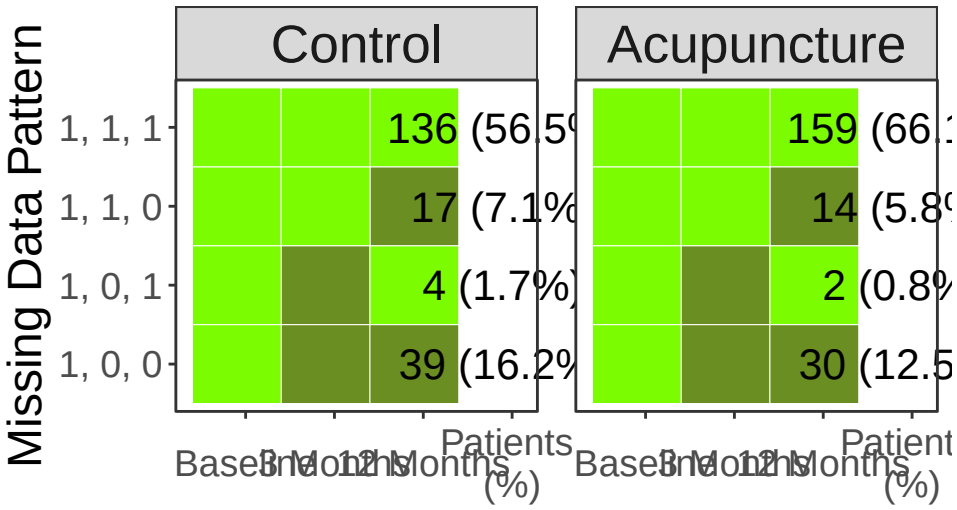


Figure 2.x: Missing data pattern is similar in both groups (1=observed, 0=missing)

VITAL data

The summary statistics of the VITAL data showed, unlike the acupuncture trial data, there are also baseline variables that contain missing data as well as the post randomised variables. A non-monotic multivariate pattern appears to show here as there is a great proportion of missing data in the knee pain score at the first year post randomisation, which drastically decreases in year 2, then steadily increases again.

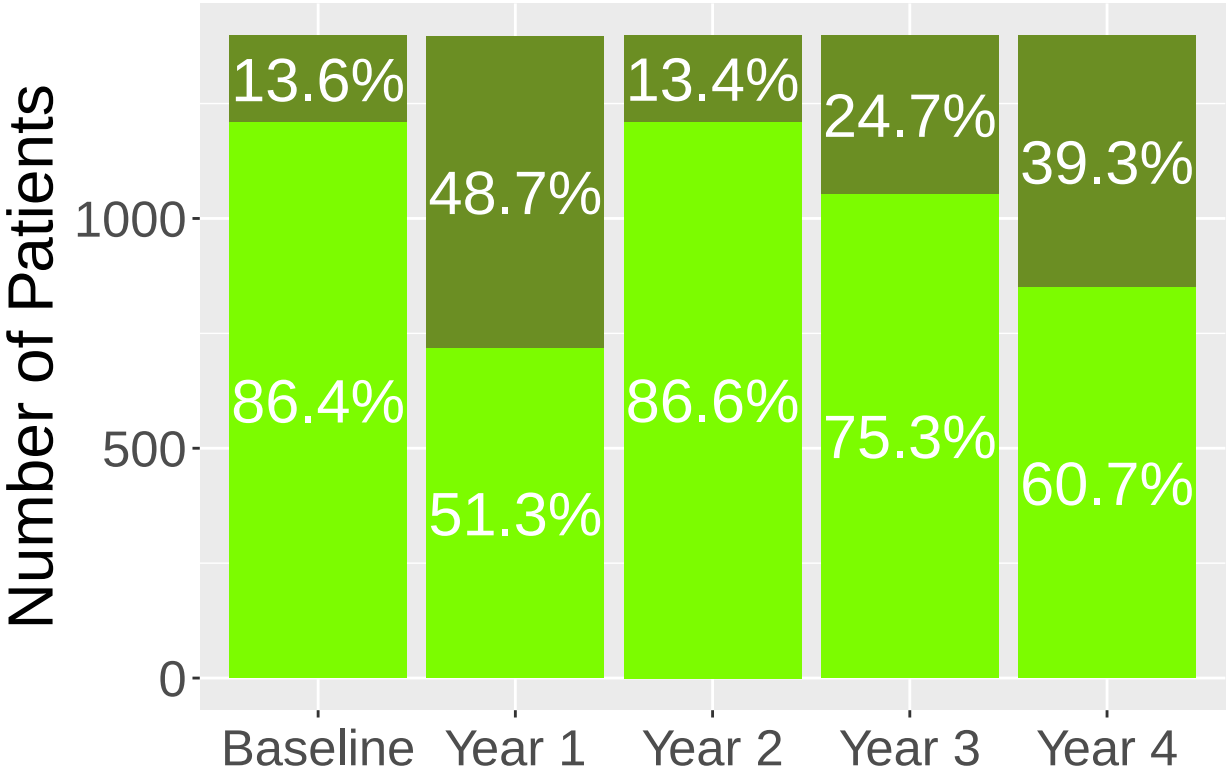


Figure 2.x: Non-monotonic pattern across out knee pain scores in VITAL data.

Recall that the VITAL data set contains 4 different combinations of treatment. Separating this out in the plot below we can see the percentage of missing data in each knee pain time point variable in each group. The proportion appears to be similar in each group with year 2 post randomisation containing the most missing data.

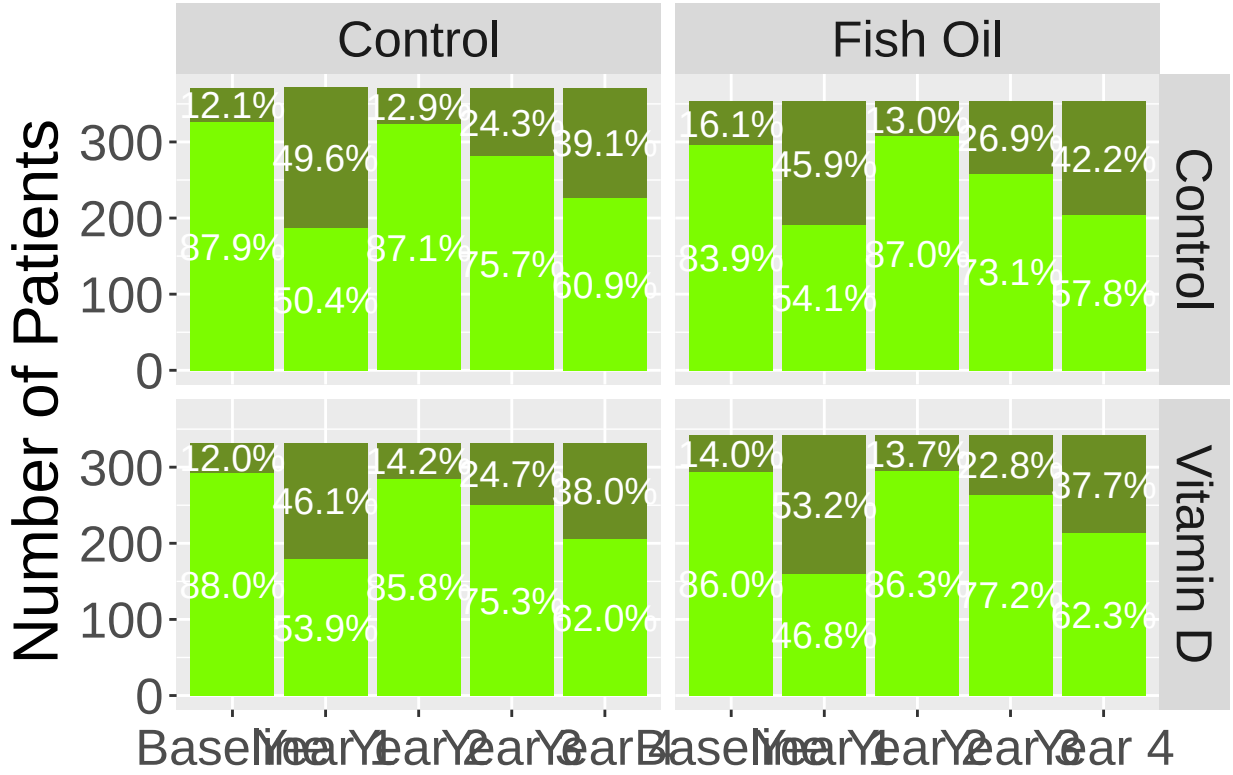


Figure 2.x: Percentage of missing data across baseline and 4 follow-up time points. Non-monotonic pattern appears similar in 4 randomised groups.

Looking at the possible missing data patterns in the 4 combinatory groups, they are mostly similar with little anomalies. As this dataset has more time-points recorded than the acupuncture trial data, and more treatmenr groups, there are more possible missing data patterns to observe. The most common data pattern still stands as (1,1,1,1,1) in which data is observed at all time-points.

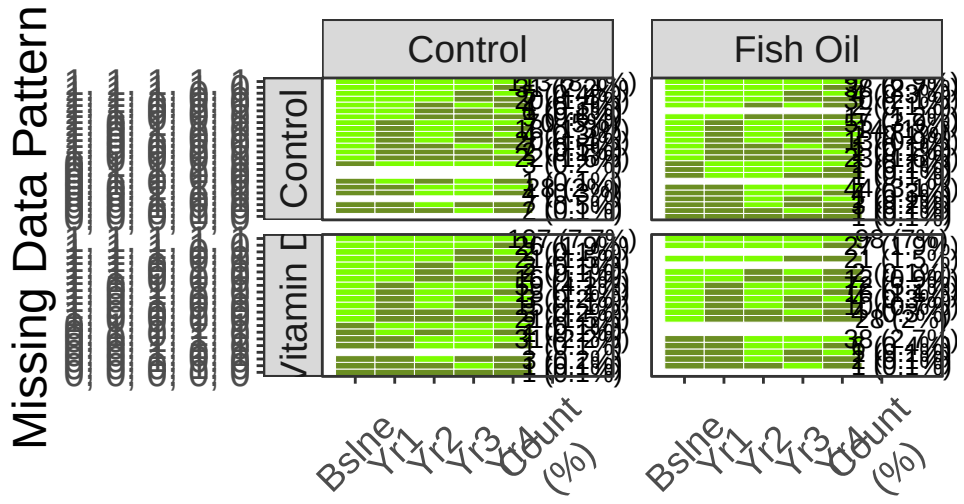


Figure 2.x: Missing data patterns across 4 treatment groups in VITAL data.

3 Methods

3.1 Data wrangling

The VITAL dataset was originally in wide format, which is not suitable for methods like linear mixed-effects models (LME) that require long-format data. To address this, the data was reshaped to long format using pivot_longer(), as illustrated in the example code provided for VITAL.

In the case of multiple imputation (MI), imputation was first performed on the wide-format data using the mice() function. The resulting imputed datasets were then combined and transformed to long format, allowing for the inclusion of time as a continuous variable. This approach enables the application of LME as the substantive model, which effectively accounts for repeated measurements over time. Moreover, it accommodates more flexible timing of follow-up assessments. An example code snippet demonstrates how the acupuncture imputed dataset was processed and analyzed using this method.

3.2 Anchoring methods

We began our analysis with basic methods to serve as a comparison benchmark for more advanced techniques. As previously mentioned, these included complete case analysis, last observation carried forward (LOCF), and the mean observation method. Each of these approaches was used to anchor the results and provide a reference point for evaluating the performance of more robust models.

3.2.1 Complete case analysis

For both data sets, given that our estimand is the mean difference in pain scores at the end of the study, we fit a model regressing the final pain score on treatment group and baseline pain score. If the data are Missing Completely At Random (MCAR), complete case analysis (CCA) can yield valid inference. For simplicity and to conserve computational resources, we did not adjust for additional baseline covariates. To maintain comparability across methods, we omit these covariates in the other analyses as well.

We applied the following models for each dataset:

One of the main drawbacks of using complete case analysis (CCA) is the loss of power due to reduced sample size, along with the risk of introducing bias. In our implementation, the lm() function in R automatically excludes any individual with missing values in the model variables. As a result, we are left with only 301 out of 401 observations in the acupuncture dataset and 698 out of 1390 in the VITAL dataset—corresponding to a loss of approximately 25% and 50% of the sample size, respectively.

Although both datasets are relatively large, the reduction in statistical power from complete case analysis may be less of a concern. However, the potential for bias remains a more serious issue. We have not included many covariates in our models, so bias could exist even with complete data due to omitted variables. Now, suppose we attempted to adjust for all relevant baseline covariates—missing data would still pose a problem. For instance, in the VITAL dataset, several baseline covariates have missing values. Excluding observations with missing data in these variables could still induce bias, considering the missingness could be related to the outcome.

In the acupuncture dataset, although there is no missingness in baseline variables, complete case analysis still has two major limitations. First, it does not allow for adjustment of interim outcomes, such as the 3-month pain score. This limitation can be addressed by using a linear mixed-effects model, which accounts for repeated measurements over time. Second, auxiliary variables—such as pain frequency—cannot be included in this model, even though they may help improve estimation. This issue can be resolved through multiple imputation, which allows the use of additional information to handle missing data more effectively.

In rare cases—such as when the sole interest is in estimating the mean difference between groups at the end of the study, and there are no important non-baseline covariates complete case analysis can still yield valid results. Thus when apply complete case analysis, we are assuming: the outcome is Missing At Random (MAR), conditional on the observed baseline predictors included in the model.

This condition sometimes referred to as "conditionally MCAR". While it is weaker than the strict MCAR assumption, it is still quite strong. It implies that:

- Missingness depends only on the variables included in the model;
- The model is correctly specified;
- The outcome variable is not involved in the missingness mechanism—consistent with the MAR framework.

Under these assumptions, CCA can provide unbiased estimates. However, they are rarely fully met in practice, especially in complex longitudinal studies where many relevant variables and time points are involved.

This can be expressed in the mathematics form, assuming outcome of interest Y MAR depends on baseline covaraites X

$$Pr(Ri = r \mid Yi, Xi) = Pr(Ri = r \mid Xi)$$

Thus, we can infer under MAR, the distribution of outcome within covaraites is the same in the observed data, the unobserved data, and the population.

$$\begin{aligned}
&Pr(Yi \mid Ri = r, Xi) \\
&= \frac{Pr(Ri = r, Yi, Xi)}{Pr(Ri = r, Xi)} \\
&= \frac{Pr(Ri = r \mid Yi, Xi)Pr(Yi, Xi)}{Pr(Ri = r \mid Xi)Pr(Xi)} \\
&Assuming \quad \Pr(Ri = r \mid Yi, Xi) = Pr(Ri = r \mid Xi) \\
&= \frac{Pr(Yi, Xi)}{Pr(Xi)} \\
&= Pr(Yi \mid Xi)
\end{aligned}$$

The key issue is that we aim to include as much relevant information as possible in the set of predictors X to make the MAR assumption more plausible and robust. As discussed above, more statistically principled methods—such as multiple imputation or mixed-effects models—allow us to incorporate a broader set of variables and handle missingness more effectively. However, it is important to note that the MAR assumption is fundamentally untestable. Therefore, to assess the robustness of our conclusions, we must perform sensitivity analyses, which are presented in later sections.

3.2.2 Last observation carry forward

Last Observation Carried Forward (LOCF) is a simple imputation method that fills in missing values based on the last observed outcome for each subject. It relies on assumptions unrelated to the missing data mechanism, meaning it does not model the reason for missingness. Instead, LOCF assumes that the outcome remains stable after dropout, which may be plausible in some clinical contexts, such as when the treatment effect has plateaued or when patients are expected to remain in a steady state.

However, this assumption is not convincing in either of the datasets analyzed here. In both the Acupuncture and VITAL trials, pain score are unlikely to be stabilized throughout, and the outcomes display trends over time, making the LOCF assumption potentially misleading and unsuitable for accurate estimation of treatment effects.

Below is the R code used to perform LOCF imputation on the data in wide format. Following imputation, we fitted the resulting dataset using the lm() function, consistent with the approach used in the complete case analysis (CAA):

3.2.3 Mean observation method

Mean imputation is another simple method that fills in missing values by replacing them with the mean of the observed data at each timepoint. Like LOCF, it is based on assumptions unrelated to the missing data mechanism and does not attempt to model why the data are missing.

This method implicitly assumes that the mean of the observed values accurately represents the population mean, which can be problematic. In fact, this can be viewed as an even more radical assumption than assuming data are Missing Completely At Random (MCAR), since it ignores potential bias introduced by the missingness and oversimplifies individual variation.

As with LOCF, this assumption is not convincing in either of the datasets analyzed here. In both the Acupuncture and VITAL trials, pain scores vary over time, and individual trajectories differ. Replacing missing values with the average observed value at each timepoint disregards this variation and may lead to underestimated variability and biased treatment effect estimates.

Below is the R code used to perform mean imputation on the data in wide format. Following imputation, we fitted the resulting dataset using the lm() function, as in the complete case analysis (CAA):

3.3 Changing imputation methods

To address missing data in a statistically principled manner, it is helpful to distinguish between two components: the imputation step and the modeling step. In our anchoring methods, we either did not perform imputation—as in complete case analysis (CCA)—or relied on single imputation techniques, such as last observation carried forward (LOCF) or mean imputation.

While methods like linear mixed-effects models (LME) can utilize most available observed data and yield unbiased estimates under the Missing At Random (MAR) assumption without imputation, they cannot incorporate auxiliary variables that are not part of the model (e.g., pain frequency in the Acupuncture dataset). In contrast, multiple imputation (MI) allows the inclusion of such auxiliary information during the imputation process.

In most realistic scenarios, doing no imputation or applying single imputation methods is likely to introduce bias and underestimate uncertainty, especially when missingness is related to unobserved outcomes. Thus, more flexible approaches—such as MI combined with appropriate modeling—are generally preferred for robust inference in the presence of missing data.

3.3.1 Multiple imputation

A key limitation of single imputation methods is that they treat imputed values as if they were observed data when fitting the substantive model. This approach fails to reflect the inherent uncertainty associated with missing values—it assumes we have perfectly recovered the unobserved data, which is never the case in practice.

In reality, the best we can do is to estimate the distribution of the missing data conditional on the observed data, under specific assumptions about the missingness mechanism (e.g., MAR). Importantly, this distribution is not adequately captured by a single draw, as is done in single imputation. Instead, to account for uncertainty,

we must generate multiple draws from this distribution, resulting in multiple imputed datasets. This is the foundation of Multiple Imputation (MI).

As we will demonstrate in the following sections, the draws used in MI must be (at least approximately) Bayesian for Rubin's variance formula to yield valid inference. By fitting the substantive model separately to each of the K completed datasets, we obtain multiple estimates that, when pooled, incorporate both within-imputation variability and between-imputation variability. This approach not only addresses bias caused by missing data but also appropriately inflates standard errors to reflect the uncertainty introduced by imputation. Rubin's rules provide a general framework for combining these estimates to obtain point estimates, variances, confidence intervals, and statistical tests.

We also briefly introduce the different methods available for performing multiple imputation. In cases with a monotonic missing data pattern, sequential regression offers a straightforward solution. For arbitrary or multivariate missingness, more flexible approaches like joint modeling or Fully Conditional Specification (FCS)—also known as multiple imputation by chained equations (MICE)—are required.

In our project, we chose the FCS approach, given its flexibility and ease of use when dealing with datasets like Acupuncture and VITAL that have complex missingness patterns across multiple variables. A brief comparison of joint modeling versus FCS is included below to justify this decision.

While we used five imputations (K = 5) for our MI procedure in this analysis, we also discuss the considerations involved in choosing the number of imputations. These include factors such as the proportion of missing data, computational cost, and the desired precision of standard errors and confidence intervals.

Rubin's rules Once multiple imputed datasets are generated, Rubin's rules are used to combine the parameter estimates and associated uncertainty across the K imputed datasets. The steps are as follows:

1. Impute K complete datasets, each containing different plausible values for the missing data. (More details in later section, we will focus on the following steps for now)
2. Fit the substantive model (e.g., a linear model or mixed-effects model) to each imputed dataset, obtaining Parameter estimate $\hat{\beta}_k$ and Variance estimate $\hat{\sigma}_k^2$ for each k=1,2,...,K
3. Compute the pooled estimate $\hat{\beta}_{MI}$ and its total variance $\hat{V}_{MI}$:

$$\hat{\beta}_{MI} = \frac{1}{K}\sum_{k=1}^{K}\hat{\beta}_k$$

$$\hat{V}_{MI} = \hat{W} + (1 + \frac{1}{K})\hat{B}$$

$$\hat{W} = \frac{1}{K}\sum_{k=1}^{K}\hat{\sigma}_k^2$$

$$\hat{B} = \frac{1}{K-1}\sum_{k=1}^{K}(\hat{\beta}_k - \hat{\beta_{MI}})^2$$

4. To test a null hypothesis $\hat{\beta}_{MI} = \beta^0$ use a t-statistic with ν degrees of freedom. This allows us to construct confidence intervals and perform hypothesis testing, reflecting the additional uncertainty due to imputation.

$$T = \frac{\hat{\beta}_{MI} - \beta^0}{\sqrt{\hat{V}_{MI}}}$$

$$\nu = (K-1)[1 + \frac{\hat{W}}{(1+1/K)\hat{B}}]^2$$

Sequential regression MI In many longitudinal studies, the missing data pattern is approximately monotonic, particularly when dropout is due to participant withdrawal, a common situation in clinical research. In such cases, later measurements are often missing while earlier ones are observed, which permits the use of sequential regression for imputation under the assumption of Missing At Random (MAR).

To justify this, consider the joint distribution of the outcome vector for individual i can be factorized as:

$$f(Y_{i,1}, Y_{i,2}, ..., Y_{i,p}) = f(Y_{i,p} \mid Y_{i,1}, ..., Y_{i,p-1}) * f(Y_{i,p-1} \mid Y_{i,1}, ..., Y_{i,p-2}) * ... * f(Y_{i,2} \mid Y_{i,1}) * f(Y_{i,1})$$

Under a monotonic missingness pattern, for each missing value $Y_{i,j}$ all preceding values $Y_{i,1}, ..., Y_{i,j-1}$ are observed. If we also assume Missing At Random (MAR), then each of these conditional distributions can be estimated directly from the observed data. This provides a principled basis for sequential regression imputation, where each variable is regressed on the variables preceding it in order, and missing values are imputed based on those conditional models.

Suppose we have $i = 1, ..., n$ individual with $j = 1, ..., p$ variables. When the missing data pattern is monotonic, sequential regression provides an efficient and valid approach to imputation under the MAR assumption. The procedure follows these steps:

1. specify the model

$$Y_{i,j} = (1, Y_{i,1}, ..., Y_{i,j-1})^T * \beta_j + e_{i,j},\ e_{i,j}^{i.i.d.} \sim N(0, \sigma_j^2)$$

2. Under the monotonic pattern, for every missing $Y_{i,j}$ we assume that $Y_{i,1}, ..., Y_{i,j-1}$ are fully observed. We fit the regression model using ordinary least squares (OLS) to obtain estimates $\hat{\beta}_j, \hat{\sigma}_j^2$
3. To incorporate uncertainty, we draw new parameters β_j^*, σ_j^{*2} from their posterior distributions:

$$\sigma_j^{*2} = \frac{\hat{\sigma}_j^2(n_j - j)}{z}$$
$$\beta^* \sim N(\hat{\beta}, \sigma_j^{*2} A_j)$$
$$A_j = (\sum_{i=1}^{n_j} x_{i,j} x_{i,j}^T)^{-1}$$

4. For individuals with missing $Y_{i,j}$ generate imputations from the model: $Y_{i,j} = (1, Y_{i,1}, ..., Y_{i,j-1})\beta^* + e_{i,j}^*,\ e_{i,j}^* \sim N(0, \sigma_j^{*2})$
5. Repeat for $Y_{i,j+1}$ until complete

This sequential regression method is applicable to datasets with a monotonic missingness pattern, such as the Acupuncture dataset, where individuals tend to drop out in a consistent, time-ordered manner. However, it is not suitable for datasets with non-monotonic missingness, such as VITAL, where some individuals may have missing values at intermediate time points but return for later follow-ups. In such cases, more flexible approaches like Joint Modeling or Fully Conditional Specification (FCS) are required to properly handle the complex, arbitrary missing data structure.

Joint modelling When the missing data pattern is non-monotonic, as in the VITAL dataset, sequential regression is no longer appropriate. Instead, we can apply joint modeling, which makes no assumption about the missingness pattern but assumes the missing data mechanism is ignorable (typically, Missing At Random).

Under joint modeling, we assume that the complete multivariate outcome vector follows a multivariate normal distribution:

$$Y \sim N(\beta,\ \Omega)$$
$$Y = (Y_{i,1}, Y_{i,2}, ..., Y_{i,p})^T$$
$$\beta = (\beta_{0,1}, \beta_{0,2}, ..., \beta_{0,p})^T$$
$$\Omega \text{ is the covariance matrix}$$

To estimate the parameters β and Ω and impute missing values, Gibbs sampling is one of the approaches. This algorithm draws each parameter in turn, conditional on the current values of all other parameters and the data.

To get priors to start Gibbs sampling. We begin with initial estimates β^0 and Ω^0, computed from the observed data. For each variable with missing values, we also generate an initial imputation Y_M^0 by sampling from the observed values of that variable with replacement. This allows us to calculate initial statistics such as $\overline{Y}^0$ and and the sample covariance matrix S^0 (Note it also used as prior sample covariance matrix S^P in each iteration)

Then, for each iteration r the following steps are performed:

1. Draw the precision matrix (inverse covariance):$\Omega^{-1,r} \sim W(n+\nu, (S_p^{-1} + S^{r-1})^{-1})$
2. Draw the mean vector:$\beta^r \sim N(\bar{Y}^{r-1}, n^{-1}\Omega^r)$
3. Impute missing values: $Y_M^r \sim f(Y_M \mid \beta^r, \Omega^r, Y_O)$
4. Update the mean and covariance estimates: $\bar{Y}^r$ the mean of the combined imputed and observed data, and S^r the sum of squares and cross-products from the combined data

After a sufficient number of burn-in iterations, we repeat this process K times to generate K imputed datasets. These are then analyzed using the substantive model of interest, and Rubin's rules are applied to pool the results, yielding valid parameter estimates and standard errors that account for the uncertainty due to missing data.

Full conditional specification Fully Conditional Specification (FCS), also known as multiple imputation by chained equations (MICE), is an extension of sequential regression imputation that relaxes the requirement that all covariate values used in the regressions be fully observed.

Importantly, when the missingness pattern is monotonic, FCS becomes equivalent to the sequential regression method discussed earlier. However, its key advantage is that it remains valid under non-monotonic missingness, making it suitable for more general data structures like the VITAL dataset.

The term "full conditional specification" refers to the fact that each variable is imputed from its full conditional distribution, given all other variables. This allows for more flexible modeling of multivariate missingness.

The general procedure involves the following steps:

1. Reorder the variables so that the overall missingness pattern is as close to monotonic as possible. This can improve stability and convergence in the imputation process.
2. Initialize missing values by filling in initial guesses—often by drawing, with replacement, from the observed values of each variable.
3. For each variable Y_j
   - Regress the observed part of Y_j on all other variables (including those with imputed values).
   - Use the fitted model to impute the missing values in Y_j, treating the other variables as given.
4. Repeat step 3 for all variables with missing data to complete one cycle.
5. Perform multiple cycles until convergence, and then repeat the entire process K times to generate K imputed datasets.

These datasets can then be analyzed with the substantive model, and the results pooled using Rubin's rules. FCS is widely used in practice due to its flexibility and implementation in tools like the mice package in R.

FCS VS joint modelling Generally, under the assumption of a multivariate normal distribution, the joint distribution uniquely determines the full set of conditional distributions, and vice versa. This means that, in theory, joint modeling and fully conditional specification (FCS) are mathematically compatible representations of the same underlying structure—provided all models are correctly specified.

In practice, however, joint modeling using a Gibbs sampler is often considered a more efficient algorithm. It also has the advantage of allowing the inclusion of prior information, which can be particularly useful when data are sparse or when integrating external knowledge into the model. Additionally, joint modeling methods can incorporate ridge parameters to stabilize the estimation of the covariance matrix, which becomes important when the number of variables is large relative to the sample size.

However, these concerns do not apply in our project, as our datasets have moderate dimensionality and sufficient sample size. Therefore, we opted to use Fully Conditional Specification (FCS) for multiple imputation.

FCS is also easier to implement, since it does not require explicit specification of a joint model or priors, and it generally requires fewer iterations to reach convergence in practical settings. For these reasons, we adopted the FCS approach using the mice package in R to perform multiple imputation throughout this project.

Finally, within the MICE framework, different imputation methods can be specified for different variable types—such as predictive mean matching or sampling from observed values. We will explore the impact of these options later in the analysis.

How to choose m Another practical consideration in multiple imputation (MI) is the choice of the number of imputations, denoted by K. While early applications often used K=5, more recent work emphasizes that the optimal number depends on the degree of missing information in the data.

A key parameter in determiningK is γ_0, which represents the fraction of missing information for the parameter of interest. Unfortunately, γ_0 is typically unknown in advance, including in our project.

To address this, Bodner (2008) proposed a simple and conservative strategy: using the proportion of complete cases in the dataset as a proxy for $1 - \gamma_0$, thereby estimating γ_0 conservatively. This approach allows for an informed yet practical choice of K, especially when precise calculation of the missing information is not feasible.

In our analysis, we used K=5 imputations as a baseline, while recognizing that more imputations may be needed in cases with higher levels of missingness or if more precise estimates of standard errors are required. We will explore the implication of changing K in the following sessions.

Choose 3-5 imputations

The classic advice for multiple imputation is to use a low number of imputations, typically between 3 and 5, when the proportion of missing information is moderate. As discussed in Rubin (1987), the argument for choosing a small K is based on the total variance estimate:$T_K = (1 + \frac{\gamma 0}{K})T_\infty$

where T_K is the variance with K imputations, T_∞ is the asymptotic variance as $K \to \infty$ and γ_0 is the fraction of missing information. Since γ_0 is typically unknown, this formula helps illustrate the trade-off between the number of imputations and efficiency.

There is often limited benefit in increasing K beyond 5. For instance, if $\gamma 0 = 30\%$, using K = 5 result in $T_m = 1.06T_\infty$, indicating only a 6% inflation in variance compared to the ideal case.

In this project, we chose to use K=5 imputations for multiple imputation, following the classical recommendation to use a low number of imputations when the proportion of missing information is moderate.

Chosse >20 imputations

While early guidelines recommended using a low number of imputations (typically K=3-5) for moderate missingness, more recent research argue that increasing K beyond this range can yield important gains in statistical efficiency.

- Royston (2004) suggested that to constrain the coefficient of variation of $ln(t_\nu\sqrt{T})$ to below 0.05—effectively keeping the width of confidence intervals within about 10% uncertainty, a minimum of K>20 is required.
- Graham (2007) argued that to achieve statistical power within 1% of the theoretical maximum, researchers should use at least K=20.
- Bodner (2008) examined how the number of imputations relates to the fraction of missing information γ_0, and its effect on p-values and confidence intervals. He recommended increasing $K = (3, 6, 12, 24, 59, 114, 258)$ for $\gamma 0 = (0.1, 0.3, 0.5, 0.7, 0.9)$ accordingly.

In some situations—such as when estimating variance components or when dealing with highly uncertain estimands—using a very high number of imputations (e.g.,K=200) may be warranted to approximate the full posterior distribution.

The main drawback of increasing K is that it leads to longer computational time. However, this is generally the only limitation, and it becomes manageable with modern computing resources. Moreover, starting with a high number of imputations gives greater flexibility: we can always test the stability or sensitivity of our results by re-analyzing a subset of the imputations (e.g., comparing the performance at K=5,10,20) without needing to re-run the entire imputation process.

Thus, while K=3-5 is often sufficient under moderate missingness and when the focus is on point estimates, using a larger K can improve robustness in more demanding settings, with minimal trade-offs beyond processing time.

$K \approx 100\lambda$

A widely cited rule of thumb proposed by White, Royston, and Wood (2011) recommends choosing the number of imputations based on the fraction of incomplete cases in the dataset, denoted as λ. Specifically, they suggest setting:$K \approx 100\lambda$

This rule has become a de facto standard, particularly in medical research, due to its simplicity and strong theoretical support. The key idea is that the number of imputations should roughly match the percentage of individuals with any missing data.

```
- The Monte Carlo error of the pooled point estimate $\hat{\beta}$ is approximately 10% of its standard
- The Monte Carlo error of the test statistic $\hat{\beta}/SE_{\hat{\beta}}$ is roughly 0.1.
- The Monte Carlo error of a p-value is approximately 0.01 when the true p-value is 0.05.
```

These error bounds are typically acceptable in applied research, ensuring stable estimates and valid inference without requiring an excessive number of imputations.

One challenge with applying this rule arises in high-dimensional settings, where the number of variables is large. In such cases, it is common for a large proportion of individuals to have at least one missing value, which can push λ close to 1. To address this, it is reasonable to use the overall missing rate (i.e., total proportion of missing cells in the dataset) as a conservative proxy for λ when needed.

This rule provides a useful upper bound for choosing K, especially when balancing the goals of statistical precision and computational efficiency.

3.4 Changing substantive model

Another important decision point in the missing data handling process is the choice of the substantive model—the model used for the final analysis after imputation. One natural alternative to standard multiple linear regression is the linear mixed-effects model (LME).

As discussed earlier, LME has the advantage of leveraging all available observed data, even in the presence of missingness. In fact, under certain conditions, LME can yield valid inferences without requiring multiple imputation, as long as the missingness mechanism is ignorable (e.g., MAR) and auxiliary variables are not essential.

Moreover, even with fully observed data—either originally complete or completed through imputation—LME remains a superior choice in many longitudinal settings because it explicitly accounts for within-subject correlation from repeated measurements. This leads to more accurate estimates and better statistical efficiency than standard linear models, which ignore the data's hierarchical structure.

In this project, we focus on multiple linear regression and LME, both of which assume a linear relationship between covariates and the outcome. However, when selecting a substantive model, it's important to evaluate this assumption. In cases where linearity is questionable, alternative models—such as polynomial regression or spline-based models—may provide a better fit and should be considered.

An additional benefit of using LME is that it allows for flexible modeling of time. Specifically, it enables us to treat time as a continuous variable, rather than as a categorical factor, which can improve interpretability and statistical power. This modeling choice will be explored further in a later section.

3.5 Examing using forest plot

To visually compare how our estimand (treatment effect) changes across different missing data methods, we present results using a series of forest plots. These plots allow us to directly assess the impact of each imputation or modeling strategy on the estimated effect and its confidence interval.

The same process was applied to both the Acupuncture and VITAL datasets. For the VITAL trial, which follows a 2×2 factorial design, we simplify the analysis—as previously discussed—by treating fish oil and vitamin D as if they were tested in two separate randomized controlled trials. This means we ignore potential interaction effects between the two treatments and estimate their effects independently, which allows for a more straightforward comparison across methods.

3.5.1 Change imputation method and substantive model

We evaluated the impact of missing data by systematically varying both the imputation method and the substantive model used in the analysis. By doing that, we try to isolate the effects of changing either the imputation strategy or the analysis model, and highlights the practical impact of each decision on the resulting treatment effect estimates.

First, as described earlier, we applied three anchoring methods: Complete Case Analysis (CAA), Last Observation Carried Forward (LOCF), and mean imputation. Each was followed by multiple linear regression to estimate the treatment effect.

Next, we performed Multiple Imputation (MI) using the predictive mean matching method with 5 imputations, again analyzing the imputed datasets using multiple linear regression—maintaining consistency with the anchoring models for fair comparison.

Then, we changed the substantive model to a linear mixed-effects model (LME), without performing any imputation. This approach uses all available observed data and accounts for repeated measurements, offering a more efficient use of the data under the MAR assumption.

Finally, we combined both improvements: we performed MI using predictive mean matching with 5 imputations, and analyzed the completed datasets using the LME model.

3.5.2 Change estimand

Assuming our primary interest remains in estimating the mean difference in pain scores at the end of the study for both datasets, we explore how this estimand can be more efficiently estimated using interim outcome data.

In the two methods that used Linear Mixed-Effects (LME) models as the substantive model in the previous section, we took advantage of the repeated measures structure by treating time as a continuous variable. This approach allows us to model pain trajectories over time, rather than focusing only on discrete timepoints. By doing so, we are able to leverage all available interim outcomes, improving the precision of the estimated treatment effect at the final timepoint—even when some observations are missing.

3.5.3 Change FSC methods

Even within the FCS (Fully Conditional Specification) framework, there are multiple imputation strategies we can explore. We begin with deterministic prediction (method="norm.predict"), which imputes missing values using predicted values from a linear regression model. Since no uncertainty is added, this is not technically multiple imputation—each missing value is replaced by a single fixed estimate.

To account for uncertainty in the outcome, we adds random noise to the regression prediction (method="norm.nob"). Further, to reflect uncertainty in the model parameters, norm draws both regression coefficients and residual variance from their posterior distributions, providing a fully Bayesian imputation (method="norm").

Alternatively, instead of model-based predictions, we can impute using observed values. predictive mean matching (method="pmm") selects donor values from observed data that are closest in predicted value to the missing case. Weighted predictive mean matching (method="midastouch") extends this by weighting the distance between predicted values, offering a more robust variant of PMM.

Or even, if we prefer a non-parametric imputation approach, we can use the methods like random sampling (method="sample"), which imputes missing values by randomly sampling from the observed values of the same variable. This method does not rely on any model or predictor variables, and instead preserves the marginal distribution of the variable.

3.5.4 Change imputation numbers

As discussed above, there are various arguments for choosing a higher number of imputations based on the fraction of missing information and the desired precision of inference. To explore the impact of increasing K, we adjusted the number of imputations accordingly in both datasets.

For the Acupuncture dataset, where approximately 25% of cases have missing data, we increased K from the default 5 to 20 and 25. For the VITAL dataset, which has around 50% missingness, we increased K from 5 to 20 and then to 50.

3.5.5 Sensitive analysis

As mentioned in our introductory section, the missingness mechanisms are unverifiable. It is most common to assess data as if they were MAR although the could fall under MNAR. We decided to conduct a sensitivity analysis as we decided it was important to investigate the robustness of the estimated treatment effect when assuming different missingness mechanism assumptions. Sensitivity analysis is a recommended assessment yet rarely conducted in practice. This could be due to overall unappreciated missing data in studies. It is also complex and time consuming to perform, which requires specialised knowledge.

Fiero et al. 2016 - systematic review of 86 cluster randomised trials, 80 reported missing outcome data, 14 trials reported sensitivity analysis

Our chosen method of sensitivity analysis is multiple imputation with δ adjustment.

- imputation model; we used a model similar to the our imputation model that we used when perfomring the missing data handling methods
- estimands categorical and cotinuous
- choosing delta

wide, impute, data to long

- code: creating an inlist of predictors for the outcome variable
- predictor matrix
- post processing; add delta values to the imputed values
- loop over delta values which represent a potential bias added to imputed values
- runs linear regression for each dataset; maxit =10 max number of iteration pools results -extracts group

other future methods

pattern mixture model: this is an alternative method of sensitivity analysis in which the distribution of the outcome is modelled by combining the distribution of the missing values and the observed values.

$$P(Y,R) = P(Y|R)P(R) = P(Y|R=1)P(R=1) + P(Y|R=0)P(R=0)$$

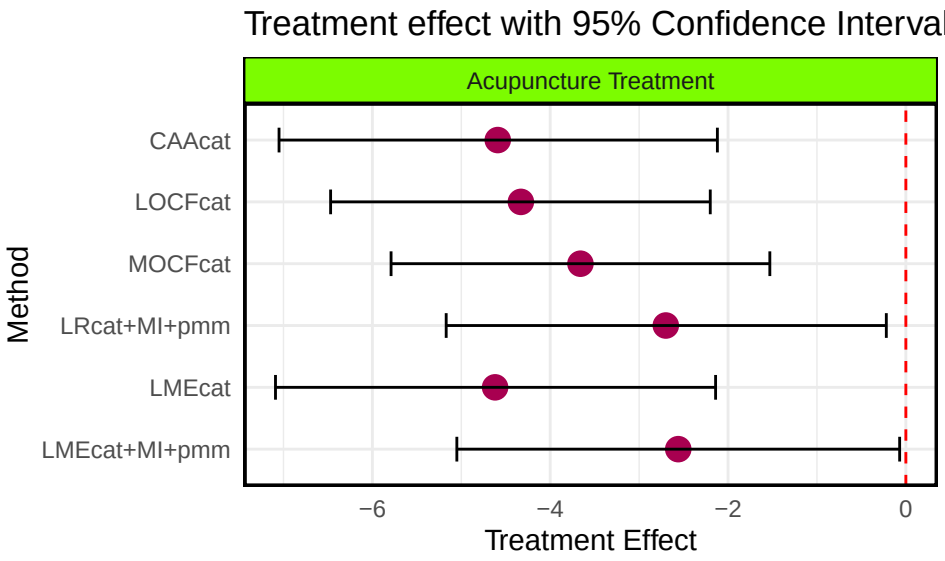
selection model; model the probability of missingness as a function of the data

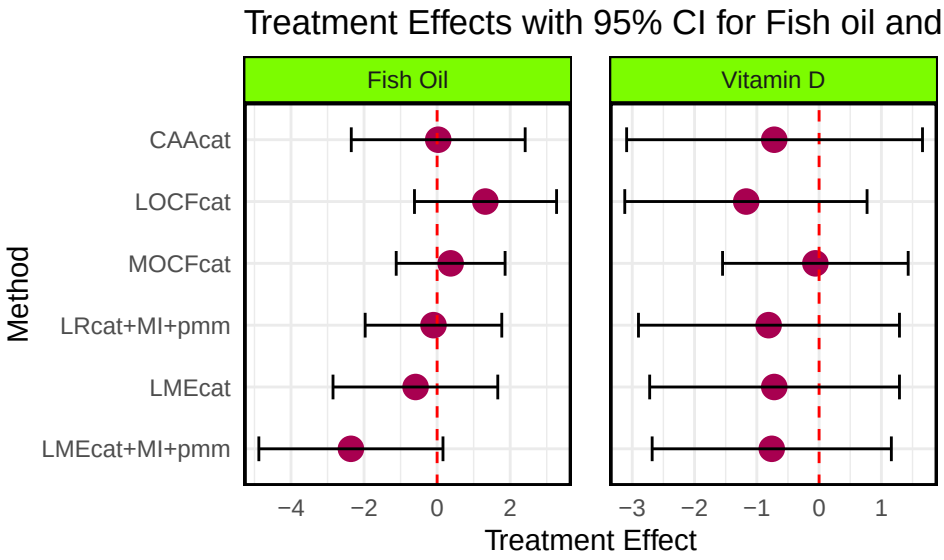
heatmap attempt for acupuncture cat_heatmap_acu_plot cont_heatmap_acu_plot

heatmap attempt for vital cont_heatmap_fishoil_plot

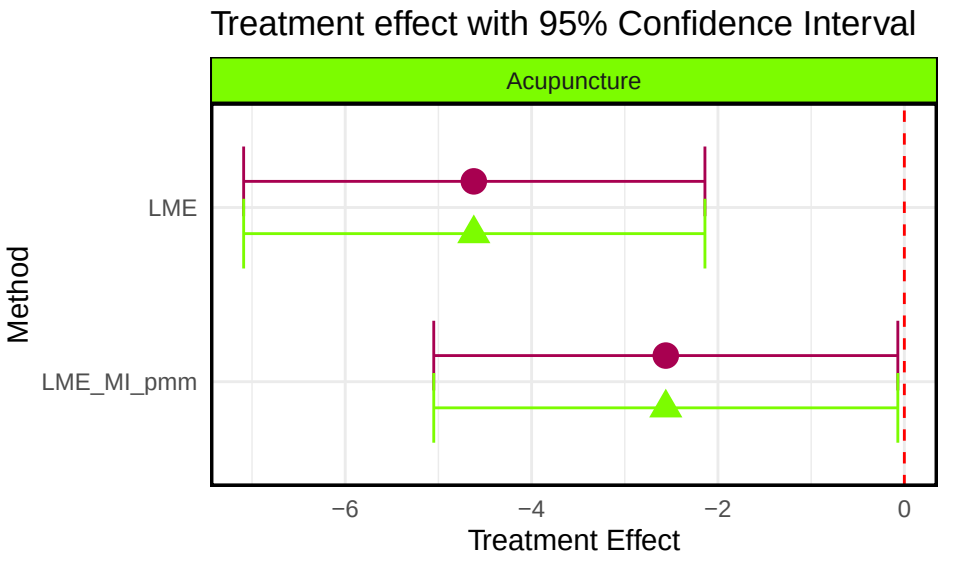
4 Result

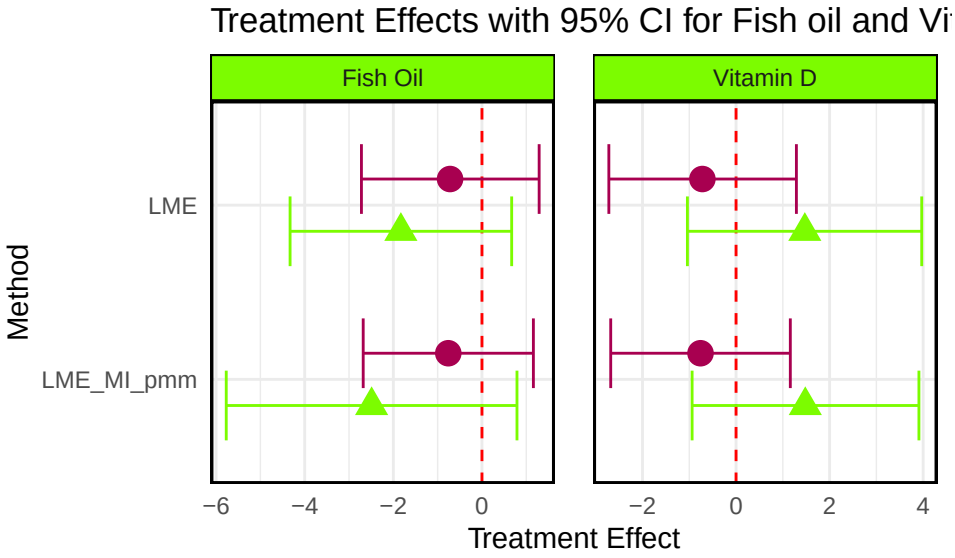
4.1 Comparing different methods



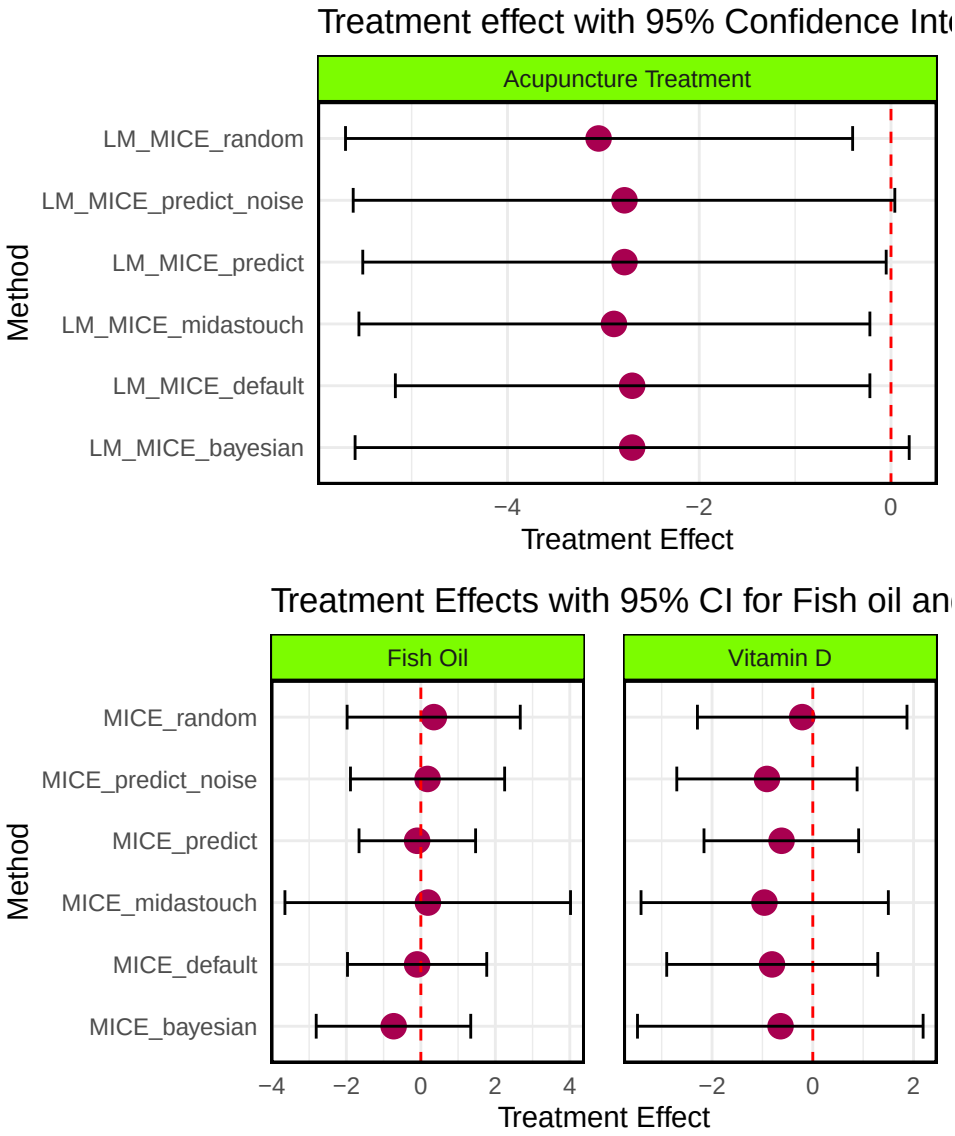


4.2 Changing estimand





4.3 Changing imputation methods and imputation numbers



4.4 Sensitive analysis

4.5 Missing data in clinical research

5 Discussion

- No meaningful difference observed
- Change estimand: Almost no impact in acupuncture study due to 2 follow-up

5.1 Limitations

- limitation: only 2 follow up for acupuncture
- limitation: VITAL only have weak therapeutic effect

5.2 Future work

- further work: simulate complete data, and compare accuracy between methods
- Change substantive model
- Use joint modelling