

LONDON
SCHOOL of
HYGIENE
& TROPICAL
MEDICINE



Cro, S (2017) Relevant Accessible Sensitivity Analysis for Clinical Trials with Missing Data. PhD thesis, London School of Hygiene & Tropical Medicine. DOI: <https://doi.org/10.17037/PUBS.03817571>

Downloaded from: <http://researchonline.lshtm.ac.uk/3817571/>

DOI: [10.17037/PUBS.03817571](https://doi.org/10.17037/PUBS.03817571)

Usage Guidelines

Please refer to usage guidelines at <http://researchonline.lshtm.ac.uk/policies.html> or alternatively contact researchonline@lshtm.ac.uk.

Available under license: <http://creativecommons.org/licenses/by-nc-nd/2.5/>

Relevant Accessible Sensitivity Analysis for Clinical Trials with Missing Data

Suzie Cro



Thesis submitted in accordance with the requirements for the degree
of Doctor of Philosophy of the University of London
January 2017

Department of Medical Statistics
Faculty of Epidemiology and Population Health
London School of Hygiene and Tropical Medicine
Funded by the MRC Clinical Trials Unit,
at University College London

Declaration

I, Suzie Cro, confirm that the work presented in this thesis is my own.

Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

The statistical analysis of longitudinal randomised controlled trials is frequently complicated by the occurrence of protocol deviations which result in incomplete datasets for analysis. However analysis is approached, an unverifiable assumption about the distribution of the unobserved post-deviation data must be made. In such circumstances it is consequently important to assess the robustness of the primary analysis of the trial to different credible assumptions about the distribution of the missing data.

Reference based multiple imputation procedures have been proposed for contextually relevant sensitivity analysis of longitudinal trials. Differences between the mean and variance of observed and missing data are specified with qualitative reference to trial arms and multiple imputation is used for estimation and inference. The primary analysis model is retained in the sensitivity analysis to assess the impact of alternative sampling behaviour on the original planned analysis. Rubin's rules are used to combine the treatment effect and variance estimates across imputed datasets, however it is unclear precisely what an appropriate measure of variance is in this setting and how Rubin's variance formula relates to this.

We begin by defining a lower bound for variance estimation in the reference based settings as the variance estimate we would obtain were we able to observe the deviation data under the postulated post-deviation data assumption. We show Rubin's variance estimate always exceeds this and moreover it approximately preserves the loss of information in the primary analysis. We also explore Rubin's variance estimate in the δ -adjusted sensitivity analysis setting and show that Rubin's variance formula preserves the loss of information in this context.

Alongside, we develop a new Stata command "mimix" for implementation of reference based sensitivity analyses. We illustrate the relevance and accessibility of the proposed methods of sensitivity analysis using data from a chronic asthma trial and a study of peer review.

Contents

Acknowledgments	13
Glossary	14
1 Introduction	16
1.1 Missing data in randomised controlled trials	16
1.2 Inference when data are missing	18
1.2.1 Mechanisms	18
1.2.2 Methods for inference	19
1.3 Sensitivity analysis with missing data	22
1.3.1 Frameworks for sensitivity analysis	22
1.3.2 Sensitivity analysis using selection models	23
1.3.3 Sensitivity analysis using pattern mixture models	24
1.3.4 Alternative approaches to sensitivity analysis	26
1.3.5 Estimands	27
1.4 Multiple imputation	28
1.4.1 Justification for Rubin’s rules	31
1.4.2 Obtaining imputations from a Bayesian posterior	34
1.5 Relevant accessible sensitivity analysis via multiple imputation	36
1.5.1 The ‘ δ -method’	37
1.5.2 Reference based sensitivity analysis	39
1.6 Two classes of sensitivity analysis	40
1.7 Motivating datasets	41
1.7.1 Peer review study	41
1.7.2 Asthma trial	42
1.8 Thesis focus	43
2 Reference based sensitivity analysis via multiple imputation	44
2.1 Reference based multiple imputation	44
2.2 Variance estimation	51
2.2.1 Motivation for evaluation	51
2.2.2 A lower bound for variance estimation	52
2.2.3 Rubin’s treatment and variance estimate	57
2.3 Exploratory simulation study	59
2.3.1 Methods	59
2.3.2 Results	60
2.3.3 Discussion	63
2.4 Summary	64

3	Behaviour of Rubin’s variance estimator in reference based multiple imputation; baseline and single follow-up	65
3.1	Information anchored sensitivity analysis	66
3.1.1	Rubin’s variance estimator under MAR	69
3.2	Rubin’s variance estimator	73
3.2.1	Copy reference	74
3.2.2	Jump to reference	80
3.2.3	Copy increments in reference	84
3.2.4	Last mean carried forward	88
3.3	Simulation study to investigate Rubin’s variance estimator	93
3.3.1	Methods	93
3.3.2	Results	94
3.3.3	Discussion	98
3.4	Baseline adjusted setting	98
3.4.1	Information anchoring for the baseline adjusted treatment estimator	99
3.4.2	Rubin’s variance estimator for the baseline adjusted treatment estimator	104
3.4.3	Simulation study to investigate Rubin’s variance estimator for the baseline adjusted treatment estimator	116
3.5	Summary	120
4	General theory for Rubin’s variance estimator in reference based analysis	121
4.1	Baseline and single follow-up setting	122
4.1.1	Setting for Proposition 1	122
4.1.2	Proposition 1	123
4.1.3	Proof of Proposition 1	123
4.1.4	Implementation for improved information anchoring	132
4.1.5	Relaxing the equal variance assumption	137
4.1.6	Extension for deviation in both arms	139
4.2	Longitudinal setting with last measured variable subject to non-response	144
4.2.1	Setting for Proposition 2	145
4.2.2	Proposition 2	145
4.2.3	Proof of Proposition 2	146
4.2.4	Implementation for improved information anchoring	149
4.2.5	Relaxing the equal variance assumption	149
4.2.6	Extension for deviation in both arms	150
4.3	Longitudinal setting with more than one deviation pattern	150
4.3.1	Setting for Proposition 3	151
4.3.2	Proposition 3	151
4.3.3	Proof of Proposition 3	152
4.3.4	Implementation for improved information anchoring	155
4.3.5	Relaxing the equal variance assumption	155
4.3.6	Extension for deviation in both arms	156
4.4	Finite imputations	160
4.5	Simulation study to explore Rubin’s variance estimator in the longitudinal setting	161
4.5.1	Methods	161
4.5.2	Results	163
4.5.3	Discussion	168
4.6	Summary	168

5	Sensitivity analysis using the ‘δ-method’	170
5.1	The ‘ δ -method’	170
5.2	Motivation for variance evaluation	172
5.3	Variance estimation in the baseline and single follow-up setting	173
5.3.1	Rubin’s variance estimator under the ‘ δ -method’	174
5.3.2	Information anchoring	178
5.4	General theory for Rubin’s variance estimator; baseline and single follow-up	179
5.4.1	Setting for Proposition 4	180
5.4.2	Proposition 4	180
5.4.3	Proof of Proposition 4	181
5.5	Simulation study; baseline and single follow-up	186
5.5.1	Methods	186
5.5.2	Results	187
5.5.3	Discussion	190
5.6	General theory for Rubin’s variance estimator; longitudinal setting	191
5.6.1	Setting for Proposition 5	191
5.6.2	Proposition 5	192
5.6.3	Proof of Proposition 5	193
5.7	Simulation study; longitudinal setting	197
5.7.1	Methods	197
5.7.2	Results	198
5.7.3	Discussion	201
5.8	Extension for deviation in both arms	201
5.9	Summary	206
6	Software and application	208
6.1	Introducing mimix	209
6.1.1	Syntax	210
6.1.2	Options	210
6.1.3	Implementation details	212
6.1.4	Specifying the imputation method	213
6.2	Sensitivity analysis of the asthma trial	213
6.3	Sensitivity analysis of the reviewer study	218
6.4	Discussion	221
7	Discussion	223
7.1	Principles for sensitivity analysis	224
7.2	Reference based sensitivity analysis via multiple imputation	224
7.3	The ‘ δ -method’	226
7.4	Generalisability	227
7.5	Information anchoring	230
7.6	Future work	231
7.7	Concluding remarks	233
	Appendices	234
A	Exploratory simulation study details	234
B	Reference based computations; baseline and single follow-up	235
B.1	The design based variance estimator when post-deviation data is observed	235

B.2	The design based variance estimator when post-deviation data is unobserved; maximum likelihood analysis result	237
B.3	Reference based imputation models	239
B.4	Imputation calculations	244
B.4.1	MAR imputation calculations	244
B.4.2	CR imputation calculations	247
B.4.3	J2R imputation calculations	251
B.4.4	CIR imputation calculations	254
B.4.5	LMCF imputation calculations	257
C	Additional results from baseline and single follow-up simulation studies	262
D	Reference based computations; longitudinal setting	273
D.1	The design based variance estimator when post-deviation data is observed	273
D.1.1	Last measured variable subject to non-response	273
D.1.2	Monotone non-response	275
E	Longitudinal simulation studies	277
E.1	Imposed missing data patterns	277
E.2	Additional results	278
F	The ‘δ-method’ computations	281
F.1	The design based variance estimator when post-deviation data is observed; baseline and single follow-up	281
F.2	The design based variance estimator when post-deviation data is unobserved; imputation calculations	282
F.3	The design based variance estimator when post-deviation data is observed; longitudinal setting	284

List of Figures

1.1	Illustration of the ‘ δ -method’	38
1.2	Illustration of jump to reference imputation	40
1.3	Asthma trial: observed mean FEV ₁ and deviation profile by treatment arm	42
2.1	MAR mean profiles in the asthma RCT	46
2.2	Drawing imputed data under MAR in the asthma RCT	46
2.3	Drawing imputed data under CIR in the asthma RCT	48
2.4	Drawing imputed data under CR in the asthma RCT	48
2.5	Drawing imputed data under LMCF in the asthma RCT	49
2.6	Rubin’s variance vs. conventional long-run sampling variance vs. lower bound for MAR, LMCF and CR	63
2.7	Rubin’s variance vs. conventional long-run sampling variance vs. lower bound for CIR and J2R	63
3.1	Rubin’s variance estimator vs. information anchored variance	95
3.2	Rubin’s variance estimator vs. information anchored variance with n=100 per arm	96
3.3	Rubin’s variance estimator vs. information anchored variance with n=1000 per arm	96
3.4	Rubin’s variance estimator vs. information anchored variance with a low covariance structure	97
3.5	Rubin’s variance estimator vs. information anchored variance with a high covariance structure	98
3.6	Rubin’s baseline adjusted variance estimator vs. the baseline adjusted information anchored variance	117
3.7	Rubin’s baseline adjusted variance estimator vs. the baseline adjusted information anchored variance with n=100 per arm	118
3.8	Rubin’s baseline adjusted variance estimator vs. the baseline adjusted information anchored variance with n=1000 per arm	118
3.9	Rubin’s baseline adjusted variance estimator vs. the baseline adjusted information anchored variance with a low covariance structure	119
3.10	Rubin’s baseline adjusted variance estimator vs. the baseline adjusted information anchored variance with a high correlation structure	119
4.1	Information anchored variance vs. RHS1 and RHS2	132
4.2	CR MI with implementation for improved information anchoring	136
4.3	J2R MI with implementation for improved information anchoring	136
4.4	CIR MI with implementation for improved information anchoring	137
4.5	Rubin’s variance estimator vs. information anchored variance in the longitudinal setting	163

4.6	Rubin's variance estimator vs. information anchored variance in the longitudinal setting with n=100 per arm	164
4.7	Rubin's variance estimator vs. information anchored variance in the longitudinal setting with n=1000 per arm	164
4.8	Rubin's variance estimator vs. information anchored variance in the longitudinal setting with a low covariance structure	165
4.9	Rubin's variance estimator vs. information preserving variance in the longitudinal setting with a high covariance structure	165
4.10	Rubin's variance estimator vs. information anchored variance in the longitudinal setting with deviation in both arms	167
4.11	Rubin's variance estimator vs. information anchored variance in the longitudinal setting with deviation in both arms; deviation active = 2x deviation placebo . . .	167
5.1	Rubin's $+\delta$ adjusted variance estimator vs. information anchored variance	188
5.2	Rubin's $-\delta$ adjusted variance estimator vs. information anchored variance	188
5.3	Rubin's $-\delta$ adjusted variance estimator vs. information anchored variance with $\sigma_{\delta}^2 = 0.01^2$	189
5.4	Rubin's $-\delta$ adjusted variance estimator vs. information anchored variance with $\sigma_{\delta}^2 = 0.05^2$	190
5.5	Rubin's $+\delta$ adjusted variance estimator vs. information anchored variance in the longitudinal setting	199
5.6	Rubin's $-\delta$ adjusted variance estimator vs. information anchored variance in the longitudinal setting	199
5.7	Rubin's $-\delta$ adjusted variance estimator vs. information anchored variance in the longitudinal setting with $\sigma_{\delta}^2 = 0.01^2$	200
5.8	Rubin's $-\delta$ adjusted variance estimator vs. information anchored variance in the longitudinal setting with $\sigma_{\delta}^2 = 0.05^2$	200
6.1	Mean FEV ₁ against time, by treatment arm, for the four different deviation (withdrawal) patterns under MAR and J2R	217
6.2	Mimix downloads	222
C.1	Rubin's variance estimator vs. information anchored variance with $\Delta = 0$	262
C.2	Rubin's variance estimator vs. information anchored variance with $\Delta = 1$	263
C.3	Rubin's variance estimator vs. information anchored variance with $\Delta = 0$ and n=100 per arm	263
C.4	Rubin's variance estimator vs. information anchored variance with $\Delta = 1$ and n=100 per arm	264
C.5	Rubin's variance estimator vs. information anchored variance with $\Delta = 0$ and n=1000 per arm	264
C.6	Rubin's variance estimator vs. information anchored variance with $\Delta = 1$ and n=1000 per arm	265
C.7	Rubin's variance estimator vs. information anchored variance with $\Delta = 0$ and a low covariance structure	265
C.8	Rubin's variance estimator vs. information anchored variance with $\Delta = 1$ and a low covariance structure	266
C.9	Rubin's variance estimator vs. information anchored variance with $\Delta = 0$ and a high covariance structure	266
C.10	Rubin's variance estimator vs. information anchored variance with $\Delta = 1$ and a high covariance structure	267

C.11 Rubin's baseline adjusted variance estimator vs. the baseline adjusted information anchored variance with $\Delta = 0$	267
C.12 Rubin's baseline adjusted variance estimator vs. the baseline adjusted information anchored variance with $\Delta = 1$	268
C.13 Rubin's baseline adjusted variance estimator vs. the baseline adjusted information anchored variance with $\Delta = 0$ and n=100 per arm	268
C.14 Rubin's baseline adjusted variance estimator vs. the baseline adjusted information anchored variance with $\Delta = 1$ and n=100 per arm	269
C.15 Rubin's baseline adjusted variance estimator vs. the baseline adjusted information anchored variance with $\Delta = 0$ and n=1000 per arm	269
C.16 Rubin's baseline adjusted variance estimator vs. the baseline adjusted information anchored variance with $\Delta = 1$ and n=1000 per arm	270
C.17 Rubin's baseline adjusted variance estimator vs. the baseline adjusted information anchored variance with $\Delta = 0$ and a low covariance structure	270
C.18 Rubin's baseline adjusted variance estimator vs. the baseline adjusted information anchored variance with $\Delta = 1$ and a low covariance structure	271
C.19 Rubin's baseline adjusted variance estimator vs. the baseline adjusted information anchored variance with $\Delta = 0$ and a high covariance structure	271
C.20 Rubin's baseline adjusted variance estimator vs. the baseline adjusted information anchored variance with $\Delta = 1$ and a high covariance structure	272
E.1 Rubin's variance estimator vs. information anchored variance in the longitudinal setting with $\Delta = 0$	278
E.2 Rubin's variance estimator vs. information anchored variance in the longitudinal setting with $\Delta = 1$	278
E.3 Rubin's variance estimator vs. information anchored variance in the longitudinal setting with deviation in both arms and n=100 per arm	279
E.4 Rubin's variance estimator vs. information anchored variance in the longitudinal setting with deviation in both arms and n=1000 per arm	279
E.5 Rubin's variance estimator vs. information anchored variance in the longitudinal setting with deviation in both arms and a low covariance structure	280
E.6 Rubin's variance estimator vs. information anchored variance in the longitudinal setting with deviation in both arms and a high covariance structure	280

List of Tables

1.1	Peer review study: quality of peer review at baseline for those who did and did not complete the second review	41
2.1	Reference based multiple imputation options	50
2.2	Performance of Rubin's Rules in a simulation study with 40% missingness	62
3.1	Difference between Rubin's CR MI variance estimator and the information anchored variance with $\rho = 0.5$ using (3.17)	80
3.2	Difference between Rubin's MI variance estimator and the information anchored variance for CR, J2R, CIR and LMCF with baseline and a single follow-up	92
3.3	The ratio of the difference between Rubin's MI variance estimator and the information anchored variance to the information anchored variance for CR, J2R, CIR and LMCF with baseline and a single follow-up	92
3.4	Difference between Rubin's MI variance estimator and the information anchored variance for CR, J2R, CIR and LMCF with baseline and a single follow-up and baseline adjustment	114
3.5	The ratio of the difference between Rubin's MI variance estimator and the information anchored variance to the information anchored variance for CR, J2R, CIR and LMCF with baseline and a single follow-up and baseline adjustment	115
5.1	Rubin's MI treatment estimate vs. true treatment effect using the ' δ -method'	187
6.1	Specifying the imputation method using mimix	213
6.2	Sensitivity analysis results for the asthma trial	218
6.3	Sensitivity analysis results for the reviewer study	221
A.1	Values of α in the model for response in the exploratory simulation study	234
B.1	Derived Rubin's variance versus simulated Rubin's variance under MAR	247
B.2	Derived Rubin's variance versus simulated Rubin's variance under MAR, for the baseline adjusted treatment effect	247
B.3	Derived Rubin's variance versus simulated Rubin's variance under CR	250
B.4	Derived Rubin's variance versus simulated Rubin's variance under CR, for the baseline adjusted treatment effect	250
B.5	Derived Rubin's variance versus simulated Rubin's variance under J2R	254
B.6	Derived Rubin's variance versus simulated Rubin's variance under J2R, for the baseline adjusted treatment effect	254
B.7	Derived Rubin's variance versus simulated Rubin's variance under CIR	257

B.8	Derived Rubin's variance versus simulated Rubin's variance under CIR, for the baseline adjusted treatment effect	257
B.9	Derived Rubin's variance versus simulated Rubin's variance under LMCF	261
B.10	Derived Rubin's variance versus simulated Rubin's variance under LMCF, for the baseline adjusted treatment effect	261
E.1	Imposed missing data patterns for deviators in longitudinal simulations 1	277
E.2	Imposed missing data patterns for deviators in longitudinal simulations 2	277
F.1	Derived Rubin's variance versus simulated Rubin's variance with $\delta_k \sim N(0.1, 0.05^2)$ adjustment	284
F.2	Derived Rubin's variance versus simulated Rubin's variance with fixed $\delta_k = 0.1$ adjustment	284

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my supervisor, Professor James Carpenter, for the invaluable guidance he has given me throughout my thesis. I thank him for being so generous with his wisdom and time. I would also like to specially thank Professor Mike Kenward for all the exceptional advice and support he has provided. Working with both James and Mike has been nothing but a pleasure and greatly inspiring.

I am also incredibly grateful to the MRC CTU at UCL, who funded my research and provided a stimulating working environment. I thank my fellow students and colleagues at the MRC CTU at UCL for making my time there so enjoyable.

Finally, I would like to thank my family and friends who have provided a great deal of encouragement and support. Especially my parents who have supported me throughout all my studies.

Glossary

Indices and symbols

$COV[a,b]$	denotes the covariance of a and b
d	indexes deviating individuals
\mathcal{D}	defines the set of deviating individuals
DF	indexes a de-facto estimate
DJ	indexes a de-jure estimate
e	residual error
$E[]$	expected value of $[]$
$full$	indexes an estimate based on full data (all planned measurements observed)
i	indexes individuals, often patients, unless defined otherwise
iid	independently identically distributed
j	indexes measurement occasions in the data set
J	total number of measurement occasions in the data set
k	indexes imputation number in the data set
K	total number of imputed data sets
m	indexes individuals with missing data
$N(a, b)$	Normal distribution with mean a and variance b
n	total number of units per treatment arm in the data set
o	indexes individuals with observed data
\mathcal{O}	defines the set of observed individuals
$O()$	describes the order of a function; $O(n)$ describes limiting behavior of order n
R_{ij}	response indicator for patient i at occasion j
V	denotes variance estimator
$VAR[]$	denotes sampling variance of $[]$
\mathbf{X}_i	vector of baseline covariates for individual i , $p \times 1$, including treatment
Y_{zij}	outcome for patient i at occasion j in treatment arm z
z	indexes treatment arm of individuals in the data set
β	regression coefficient
$\boldsymbol{\beta}$	column vector of regression coefficients
ϕ	generic vector-parameter for the missing data mechanism
$\boldsymbol{\theta}$	generic vector-parameter, typically $p \times 1$
θ_k	scalar treatment effect of interest in imputed data set k
θ_{MI}	scalar treatment effect averaged over the K imputed data sets
χ_v^2	Chi-squared distribution with v degrees of freedom

Abbreviations

CC	Complete Case
CI	Confidence Interval
CIR	Copy Increments in Reference
CR	Copy Reference
EM	Expectation Maximisation
EMA	European Medicines Agency
FCS	Fully Conditional Specification
FDA	Food and Drug Administration
FEV ₁	Forced Expiratory Volume in 1 second (measured in Litres)
GLM	Generalized Linear Model
IPW	Inverse Probability Weighting
ITT	Intention To Treat
J2R	Jump to Reference
JSW	Joint Space Width
L	Litre
LHS	Left Hand Side
LMCF	Last Mean Carried Forward
LOCF	Last Observation Carried Forward
MAR	Missing At Random
MCAR	Missing Completely At Random
MCMC	Markov Chain Monte Carlo
MCSE	Monte Carlo Standard Error
MI	Multitple Imputation
MICE	Multiple Imputation by Chained Equations
ML	Maximum likelihood
MMRM	Mixed Model for Repeated Measures
MNAR	Missing Not At Random
MVN	Multivariate Normal
RCT	Randomised Controlled Trial
RHS	Right Hand Side
RQI	Review Quality Index (range 1 to 5)
SD	Standard Deviation
SE	Standard Error

Chapter 1

Introduction

1.1 Missing data in randomised controlled trials

Longitudinal randomised controlled trials (RCTs) are widely used in medical research and provide essential evidence for the evaluation of new and existing treatments and interventions. Unfortunately protocol deviations, such as treatment withdrawal, partial compliance or loss to follow-up are unavoidable during the full course of a trial. Consequently we often cannot measure what we intended to for deviating individuals. Planned outcomes may be unobtainable due to the type of the deviation and, depending on the nature of the analysis, even values that were recorded post-deviation may be best regarded as missing. The result is a missing data problem, complicating the analysis.

Complexity arises with missing data because, in any analysis, we are forced to make an assumption about the distribution of the unobserved data. If the wrong assumption is made the obtained treatment effect and its standard error will be biased, resulting in misleading inferences. This could have clinically disastrous implications for patients.

Crucially the missing data assumption cannot be empirically verified from the observed data, thus in the presence of missing data it can never be fully ascertained that the resulting treatment effect is unbiased. To understand how far the key inferences depend on the missing data assumption, analysis of incomplete data should therefore consist not only of a primary analysis, under the most plausible missing data assumption, but include alternative analyses, which make a range of different credible assumptions for the unobserved post-deviation data; that is include sensitivity analysis [1].

There are many forms of sensitivity analysis [2, 3]. We define sensitivity analysis, using the definition provided by Daniels and Hogan [4] as an,

“assessment of sensitivity of model-based inferences to assumptions that cannot be verified or checked within the data.”

In the current missing data setting we are specifically interested in variation in model-based in-

ferences with regards to untestable assumptions about the distribution of the unobserved data. Ideally, inferences will be stable across sensitivity analysis indicating the missing data does not seriously affect the interpretation of results. But, it is even more important if this is not the case since it allows trialists to assess under what conditions results change and how plausible these conditions are.

The endemic of missing values in the clinical trial arena was recently highlighted by Bell et al. [5]. Out of the 77 published RCTs they reviewed in four leading medical journals between July and December 2013, 73 (95%) had partly missing outcome data. Despite an abundance of statistical methods for handling incomplete data and sensitivity analysis [6, 7, 8] only 18 (25%) of the trials with missing data conducted sensitivity analysis that altered the missing data assumption made in the primary analysis. An earlier comparable review of 71 RCTs published in the same journals between July and December 2001 by Wood et al. [9] identified only 13 (21%) trials with missing data reported a sensitivity analysis, illustrating how not much has changed during this period.

This is worrying since substantially different results can be obtained in sensitivity analysis. For example, in a RCT comparing two interventions (brief negotiation or direct advice) against a control for increasing physical activity Wood et al. [10] found if dropouts were assumed to have similar activity levels to those observed in their randomised intervention arm, i.e. brief negotiation dropouts were assumed to behave like those observed in the brief negotiation arm and direct advice dropouts were assumed to behave like those observed in the direct advice arm, then the brief negotiation intervention was significantly more effective than the direct advice intervention. If all dropouts were alternatively assumed to behave like those in the control group the effect fell below significance.

Recent regulatory guidelines from the European Medicines Agency (EMA) [11] and a Food and Drug Administration (FDA) mandated panel report from the US National Research Council [12] emphasise the importance of conducting sensitivity analysis in this context. Both reports also highlight a need for accessible and relevant methods of sensitivity analysis, where the changes in assumptions are directly applicable to the primary analysis.

Controlled imputation procedures have been proposed for contextually relevant sensitivity analysis of longitudinal trials with protocol deviation [7, 13]. Trialists are currently using such methods (see for example [14] and [15]) and their use is gaining in popularity [16]. However there is uncertainty about the appropriate variance estimator for the treatment estimator when these techniques are used [17]. We expand upon this issue in detail in the following chapter. The overall goal of this research is to evaluate these procedures and establish the required variance estimator within the context of sensitivity analysis.

We begin this first chapter with an outline of the well established typology for the mechanisms generating missing data and discuss appropriate methods for inference for each class of missingness. We then outline frameworks to investigate the sensitivity of inferences to missing data in the trial setting and introduce the controlled imputation techniques that form the main focus of this thesis.

1.2 Inference when data are missing

First we introduce notation and terminology that is used throughout this and subsequent chapters. Consider a typical longitudinal RCT and let Y_{ij} denote the planned response measurement for each patient i at time j where $i = 1, \dots, N$ and $j = 1, \dots, J$. For each patient, planned measurements can be grouped into the vector $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iJ})$. Independence between patients is assumed. The distribution of the measurement data, for patient i , is defined $f(\mathbf{Y}_i | \mathbf{X}_i, \boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is the key vector-parameter of interest and \mathbf{X}_i is the vector of observed covariates, including treatment, for patient i .

Missing data are defined as values that are not available but would be meaningful for analysis if collected [12]. When data that we intend to collect are missing, the reasons for the data being missing play a key role in informing the analysis. That is, the selection or missing data mechanism operating. We denote the response indicator as $R_{ij} = 1$ if Y_{ij} is observed, or $R_{ij} = 0$ if Y_{ij} is unobserved. For each patient, missing data indicators can be grouped into the vector $\mathbf{R}_i = (R_{i1}, \dots, R_{iJ})$. The missing data mechanism is formally defined as the vector process generating \mathbf{R}_i and is modelled as the conditional distribution, $f(\mathbf{R}_i | \mathbf{Y}_i, \mathbf{X}_i, \boldsymbol{\phi})$ where $\boldsymbol{\phi}$ is the vector-parameter for the missing data mechanism. $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ are separate/distinct.

In the presence of dropout the measurement process and missingness mechanism must be considered simultaneously as, $f(\mathbf{Y}_i, \mathbf{R}_i | \mathbf{X}_i, \boldsymbol{\theta}, \boldsymbol{\phi})$. \mathbf{Y}_i can be partitioned into two sub-vectors, \mathbf{Y}_{io} and \mathbf{Y}_{im} , where \mathbf{Y}_{io} is the sub-vector of observed responses ($R_{ij} = 1$) and \mathbf{Y}_{im} is the sub-vector of missing responses ($R_{ij} = 0$).

If the measurements can be ordered in such a way that, for a patient i , Y_{ij} missing implies Y_{ij^*} is missing for all $j^* > j$ and responses $1, \dots, j - 1$ are observed then the missing data pattern is referred to as monotone. Dropout in a longitudinal trial is an example of a monotone missingness pattern. Alternatively the missing data pattern is non-monotone.

1.2.1 Mechanisms

There are three broad classes of missing data mechanisms originally introduced by Rubin in 1976 [18], that predicate the statistical handling of missing data: Missing Completely At Random (MCAR), Missing At Random (MAR) and Missing Not At Random (MNAR).

A missingness process is said to be MCAR where the probability of data being missing does not depend on the unobserved values of the data themselves, or the observed values of other recorded variables. More broadly, missingness is unrelated to the inference we wish to draw. That is, $f(\mathbf{R}_i | \mathbf{Y}_i, \mathbf{X}_i, \boldsymbol{\phi}) = f(\mathbf{R}_i | \boldsymbol{\phi})$. Consider a trial comparing vitamin D supplementation against placebo for progression of knee osteoarthritis. The primary outcome measure is Joint Space Width (JSW) in the knee measured from an X-ray at baseline, one year and three years. If the X-ray machine is not working on a particular day, all patients who attend the clinic that day for follow-up will not be able to obtain an X-ray and their JSW measure will consequently be missing for that visit. In this case the data will be MCAR as the missingness does not depend on the underlying JSW measurement or any patient characteristics, it is unrelated to the inferences we wish to draw.

In the less strict case of MAR, the probability of data being missing does not depend on the unobserved values of the data themselves, given observed information. That is the missingness depends on observed values marginally, but given the observed data is conditionally independent of the missing data, i.e. $f(\mathbf{R}_i|\mathbf{Y}_i, \mathbf{X}_i, \phi) = f(\mathbf{R}_i|\mathbf{Y}_{i0}, \mathbf{X}_i, \phi)$. In our example missingness at three years will be MAR if given the observed data (baseline, one year JSW and treatment group) the unseen response provides no additional information on the reason for non-response.

If in addition to MAR (or MCAR) the parameters of the missingness process (ϕ) and those of the data measurement process (θ) are distinct, such that the joint parameter space is the product of the two separate parameter spaces, then the missing data mechanism is termed ignorable.

The missingness process is termed MNAR where even given the observed data the probability of data being missing does depend on the unobserved values of the data themselves. That is, $f(\mathbf{R}_i|\mathbf{Y}_i, \mathbf{X}_i, \phi) \neq f(\mathbf{R}_i|\mathbf{Y}_{i0}, \mathbf{X}_i, \phi)$. This is also often referred to as non-ignorable missingness. In our example missingness at three years will be MNAR if a patient experiences a sudden decline in JSW sometime after the year 1 visit e.g. due to a fall, which greatly impacts their mobility and renders them unable to attend the planned follow-up. The observed data up to time 1 does not capture the reason for non-response.

There is no test that will definitively reveal which missingness mechanism is operating within any dataset. Although MCAR can be distinguished from MAR, e.g. via a logistic regression of observed outcomes and/or covariates on missingness, the data at hand cannot confirm which mechanism is operating. Since we can never know what the missing values are, we cannot distinguish between MNAR and both MAR and MCAR. Thus sensitivity analysis, in the sense introduced above, plays an important role. In collaboration with the trial team/regulators we must pick the most plausible assumption for the data at hand, conduct primary analysis under that assumption and then perform sensitivity analysis under alternative plausible missing data assumptions to assess the robustness of the results.

1.2.2 Methods for inference

The process of making assumptions is separate, but informs the statistical methods we use for parameter estimation and inference with missing data. When taking a frequentist approach to inference, two issues arise when we use missing data techniques. These are finding unbiased parameter estimates (that home in on the true value as the sample size increases i.e. are consistent) and providing variance estimators or standard errors that are reliable for inferential purposes.

Within this thesis the *sampling variance* of an estimator is defined as the variance of the estimator of interest over repeated sampling from the population, i.e. the variance in a long-run sense over an assumed data mechanism. The *estimated variance* is the variance of the estimator of interest computed using the analysis model from the single sample of data available to the analyst. When the assumptions of the analysts modelling procedure correspond with the assumptions for the data generating mechanism the sampling variance and estimated variance will asymptotically agree.

The most basic, but common, Complete Case (CC) analysis approach ignores patients with missing data and includes only completely observed cases in the analysis i.e. patients who complete the trial. This approach provides unbiased results under MCAR and MAR provided the analysis

model includes treatment and all covariates associated with the missingness. For example when the analysis consists of a linear regression model relating an outcome Y to one or more predictors X , if missingness occurs in the outcome Y or one or more of the predictors X (or both Y and X), fitting the regression model to the completers will be unbiased provided the probability of being a completer is independent of Y conditional on X . But due to a reduced sample size results will be less precise than when full data is observed, an inevitable consequence of missing data. We can immediately see this in a longitudinal trial setting where the unadjusted mean of the outcomes at follow-up time point J , $Y_{i,J}$, over all patients is of interest, where we assume $Y_{i,J} \sim N(\mu_J, \sigma^2)$. If $Y_{i,J}$ is observed for all N cases, the estimated variance of the mean in expectation will be σ^2/N where σ^2 denotes the variance of $Y_{i,J}$. But if $Y_{i,J}$ is observed for only a subset of n_o cases, the estimated variance of the expected mean from CC analysis will be σ^2/n_o in expectation. The CC estimated variance is greater by a factor of N/n_o . So while CC analysis has its advantages in simplicity, it can be an inefficient method especially if there are post-randomisation pre-primary end point data.

Under ignorability (union of MAR and MCAR) a variety of options, including Likelihood-based methods or Bayesian inference, which are based on all the observed data, will provide valid inference for the parameter of interest (θ). To see this, we show that the contribution for patient i to the observed data likelihood is obtained by integrating out the missing data as follows,

$$f(\mathbf{Y}_{io}, \mathbf{R}_i | \mathbf{X}_i, \theta, \phi) = \int f(\mathbf{Y}_{io}, \mathbf{Y}_{im} | \mathbf{X}_i, \theta) f(\mathbf{R}_i | \mathbf{Y}_{io}, \mathbf{Y}_{im}, \phi) d\mathbf{Y}_{im}.$$

Since under MAR, $f(\mathbf{R}_i | \mathbf{Y}_{io}, \mathbf{Y}_{im}, \phi) = f(\mathbf{R}_i | \mathbf{Y}_{io}, \phi)$, the observed likelihood contribution for patient i can be re-written as,

$$f(\mathbf{Y}_{io}, \mathbf{R}_i | \mathbf{X}_i, \theta, \phi) = \int f(\mathbf{Y}_{io}, \mathbf{Y}_{im} | \mathbf{X}_i, \theta) f(\mathbf{R}_i | \mathbf{Y}_{io}, \phi) d\mathbf{Y}_{im},$$

which simplifies to,

$$\begin{aligned} f(\mathbf{Y}_{io}, \mathbf{R}_i | \mathbf{X}_i, \theta, \phi) &= f(\mathbf{R}_i | \mathbf{Y}_{io}, \phi) \int f(\mathbf{Y}_{io}, \mathbf{Y}_{im} | \mathbf{X}_i, \theta) d\mathbf{Y}_{im} \\ &= f(\mathbf{R}_i | \mathbf{Y}_{io}, \phi) f(\mathbf{Y}_{io} | \mathbf{X}_i, \theta). \end{aligned}$$

Thus under ignorability and further when θ and ϕ are disjoint, valid inferences can be obtained from the observed likelihood only. For example, the analysis of a longitudinal trial under ignorability can be appropriately conducted using a mixed model for repeated measures (MMRM).

Alternatively, single imputation methods may be employed for inference, which entail substituting a reasonable ‘guess’ for the missing data and analysing the data as if it were complete. Many different imputation methods can be employed from simply substituting the mean of the observed data on the same variable (unconditional mean imputation) to conditional mean imputation, which uses

a regression model based on observed variables to predict missing values. Consider a longitudinal trial with $J = 2$ time points with planned measurements Y_{i1} and Y_{i2} , where Y_{i1} is observed for all patients however Y_{i2} is only partially observed. Conditional mean imputation is achieved by first regressing Y_{i2} on Y_{i1} in the observed data. The regression parameter estimates are then used to predict Y_{i2} for patients missing this outcome as follows,

$$Y_{i2} = \hat{\beta}_0 + \hat{\beta}_1 Y_{i1}. \quad (1.1)$$

Simple imputation methods however suffer from a key downfall. In analysing the imputed data as if it were real data, standard errors are underestimated and test statistics overestimated. In the case of unconditional mean imputation the estimated variance of the observed and imputed values is $s_o^2(n_o - 1)/(N - 1)$, where s_o^2 is the estimated variance from the n_o observed cases. Since s_o^2 is a consistent estimator for the true sampling variance under MCAR, the estimated variance underestimates the variance by a factor of $(n_o - 1)/(N - 1)$.

An improvement upon both simple imputation methods is stochastic regression imputation which incorporates an element of randomness into the process. In our example, instead of imputing from (1.1) we impute from,

$$Y_{i2} = \hat{\beta}_0 + \hat{\beta}_1 Y_{i1} + e_i,$$

where $e_i \stackrel{iid}{\sim} N(0, \hat{\sigma}_{2.1})$ and $\hat{\sigma}_{2.1}$ is the residual variance from the regression of observed Y_{i2} on Y_{i1} . However, this still does not fully take into account the uncertainty of imputation. The imputed values are given the same status as observed values and we haven't acknowledged the uncertainty in estimating β_0, β_1 and $\sigma_{2.1}$.

Multiple Imputation (MI) [19] is a popular technique that addresses this downfall by repeating the imputation more than once to provide a valid estimate of the standard error of the parameter estimates. Random draws are taken using an appropriate Bayesian predictive distribution and the same analysis that would be undertaken had the data been complete is fitted to each imputed dataset. A set of combination rules are used to provide one overall estimate and an estimate of precision. The variability across the multiple imputations is used to adjust the standard error appropriately upwards to reflect the imputation uncertainty. The MI procedure is very flexible and outlined in full detail in Section 1.4.

An alternative approach to analysis under MAR weights the observed data by the probability of non-response. The probability of non-response can be estimated as a function of the observed responses e.g. via a logistic regression of the response indicator on observed covariates and predictors of response. The (analysis) model relating the response to the explanatory variables and covariates is then fit using weights which are based on the observed probability of non-response. This is referred to as inverse probability weighting (IPW) [20, 21].

Under MNAR, because the missingness mechanism cannot be factored out from the joint distribu-

tion of the data and the missingness mechanism, inference must be based on the joint distribution. That is we must postulate a model for both the missingness and data. Fitting MNAR models, as described in [4] and [8], can be more computationally demanding. But since we will never know the exact missing data mechanism, MNAR modelling—which we expand upon below—most often cannot be ruled out.

1.3 Sensitivity analysis with missing data

In any incomplete data setting there is not one analysis that can be considered definitive. Since unverifiable assumptions are required for the analysis, we should postulate the most plausible missingness mechanism and perform a valid analysis for that class of missingness for the trials primary analysis. Subsequently we should postulate alternative plausible missingness mechanisms and perform valid analysis as sensitivity analysis.

1.3.1 Frameworks for sensitivity analysis

In the clinical trial setting it is recommended that ignorable (MAR) likelihood based methods be used for primary analysis [6, 7, 8]. The strong MCAR assumption is unlikely to be valid, particularly in longitudinal settings when data are missing due to uncontrollable events, since these events are often associated with the study variables. Analysis under MCAR is additionally very inefficient in this setting. Although not verifiable, MAR can often be considered the most plausible assumption. Further analysis under MAR, unlike MNAR, does not require the modelling of the dropout procedure. Any analysis under a postulated MNAR assumption is heavily assumption led thus cannot be considered definitive. Since we can never be sure of the model for the dropout MAR is a natural starting point. This is the approach we adopt to primary analysis throughout this thesis.

Since analysis under the assumption of MAR generally forms the primary analysis, procedures for sensitivity analysis focus on approaches to MNAR analysis where the measurement process and missingness mechanism must be jointly modelled. As highlighted by Little and Rubin [22] the joint distribution can be factorised in two different ways,

$$f(\mathbf{R}_i|\mathbf{Y}_i, \mathbf{X}_i, \phi) f(\mathbf{Y}_i|\mathbf{X}_i, \theta) = f(\mathbf{Y}_i, \mathbf{R}_i|\mathbf{X}_i, \theta, \phi) = f(\mathbf{Y}_i|\mathbf{R}_i, \mathbf{X}_i, \theta) f(\mathbf{R}_i|\mathbf{X}_i, \phi)$$

Either a model for the missing data mechanism, \mathbf{R}_i given the measurement data, with a marginal model for the data \mathbf{Y}_i as expressed on the left can be specified or the conditional distributions of the response data given the fully observed data on the right with a marginal model for the missingness process can be given. The former factorisation is referred to as the selection model, whilst the latter is termed the pattern mixture model.

There are many ways in which either type of model can then be fully specified, however many specifications will be practically implausible. In the RCT setting we desire a principled accessible

framework for approaching this. A commonly advocated way to perform sensitivity analysis is to explore departures from the joint distribution implied by MAR [4, 7]. MAR provides an unambiguous starting point for MNAR exploration. In Section 1.2 the selection form of the MAR assumption was presented. The MAR assumption can be expressed in the pattern mixture form as, $f(\mathbf{Y}_i, \mathbf{X}_i | \mathbf{R}_i, \boldsymbol{\theta}) = f(\mathbf{Y}_{im} | \mathbf{Y}_{io}, \mathbf{X}_i, \boldsymbol{\theta})$. This reveals MAR implies the conditional distributions of later response data given earlier response data are equal for patients who do and do not complete. Molenberghs et al. [23] show that this relationship holds quite generally and each selection model form implies a pattern mixture form. It naturally follows that for sensitivity analysis one can model departures from MAR either via the selection process or within the different patterns of data.

1.3.2 Sensitivity analysis using selection models

Selection models consist of a multivariate model for the response and a model for the reason for missing data. The precise form for each model depends on the specific trial context. The two models are fitted together. The models can be fitted together in a frequentist framework [7, 24, 25]. Numerical integration over the missing data is required. Since numerical convergence problems may be experienced model fitting may alternatively be done using MCMC methods, which can be implemented in `winBUGS` [7, 26].

Diggle and Kenward [24] describe a general selection modelling approach for dropout in longitudinal studies consisting of a multivariate linear model for the underlying measurement response, with a logistic discrete hazard model for the selection process. These two models are then fitted together. Consider a simple trial with a single follow-up Y_i for each patient i . The probability of dropout is related to the possibly unobserved outcome and any baseline covariates, \mathbf{X}_i , via a linear logistic link,

$$\text{logitPr}(R_i) = \alpha + \mathbf{X}_i + \delta Y_i.$$

The log odds of response depend on the baseline covariates \mathbf{X}_i and the response Y_i for non-zero values of δ . Departures from MAR are consequently expressed in terms of differences in the (log) odds of response per unit change in the response, conditional on all the other variables in the model — as δ —.

For sensitivity analysis, a series of selection models can be run using plausible values of δ to investigate how results change under alternative credible MNAR missing data assumptions. Plausible values should be identified through discussions with clinical experts on the trial team. When $\delta = 0$, this model represents analysis under MAR. Thus larger δ represent greater departures from MAR.

Specifying differences in the odds of response per unit change in the unobserved outcome given patients characteristics (δ), may however not seem natural or even be understandable to all clinical colleagues. Therefore such models may not be immediately usable. For longitudinal trials we will need to incorporate additional sensitivity analysis parameters (δ) in the model for the selection process to reflect dependence of missingness on previously measured outcomes, which can also

differ by treatment arm and patients characteristics, further complicating the interpretation.

1.3.3 Sensitivity analysis using pattern mixture models

For sensitivity analysis to be useful, methods and assumptions must be transparent and interpretable to all involved in the trial, not just the experienced statistician. If the assumptions are not understood then the results cannot be. A pattern mixture model requires specification of the joint distribution of the partially and fully observed response variables \mathbf{Y}_i , for each pattern of missing data, which implies the conditional distribution of partially observed response data given the observed response data within each pattern as follows,

$$f(\mathbf{Y}_i|\mathbf{R}_i, \mathbf{X}_i, \boldsymbol{\theta}) f(\mathbf{R}_i|\mathbf{X}_i, \boldsymbol{\phi}) = f(\mathbf{Y}_{im}|\mathbf{Y}_{io}, \mathbf{R}_i, \mathbf{X}_i, \boldsymbol{\theta}) f(\mathbf{Y}_{io}|\mathbf{R}_i, \mathbf{X}_i, \boldsymbol{\theta}) f(\mathbf{R}_i|\mathbf{X}_i, \boldsymbol{\phi}).$$

In the pattern mixture framework, the assumptions correspond directly to what is observed i.e. the data in the different subgroups of the trial having different distributions. So the pattern mixture approach is a convenient way to proceed. By construction pattern mixture models are underidentified [27]. The observed data does not reveal the distribution of the unobserved data. Various identifying restrictions have been discussed in the literature which can be used for analysis. These include so-called complete case missing values (CCMV) and neighbouring case missing values (NCMV), where identification is done based on either the completers' or neighbours' pattern [27]. Identification via setting all distributions that condition on the same set of observations as equal, known as available case missing values (ACMV), was also proposed in [27]. ACMV corresponds to MAR in the pattern mixture framework [23].

The distribution of the observed data i.e. $f(\mathbf{Y}|\mathbf{X}, \mathbf{R} = 1)$ provides a natural starting point to explore departures from. MNAR assumptions can then be framed in a transparent manner, with explicit description of how data for completers differs to incompleters. We therefore argue, in line with White et al. [28] and Daniels and Hogan [4] that the pattern mixture approach lends itself to more accessible assumptions. Scharfstein et al. [29] argue that parameters governing dropout in a selection model framework are more plausibly determined a-priori. However we agree with White et al. [28] who dispute the relevance of this point if non-informative priors are incorporated into the pattern mixture framework.

As discussed by Daniels and Hogan [4], starting with specification of the conditional data distribution implied by MAR, one can readily perform sensitivity analysis exploring departures from MAR by utilising location and scale shifts to create shifted distributions for the unobserved measures. Trial statisticians should seek clinical experts opinion on the degree of departure from MAR, e.g. for a single continuous outcome the anticipated likely mean difference in outcome between completers and non-completers. This can be directly incorporated into the analysis using a pattern mixture approach.

For example, discussions with clinical colleagues can help to identify by how much on average the unobserved deviators outcomes are expected to be worse or better than those observed. In a depression trial it could be specified that the observed rate of decline in a depression score for those

that deviate is for example, 20%, 50% or 75% worse than that observed for the patients in the trial. Alternatively that the outcome is worse than for those observed in the trial at a specific time point by 0.25SD, 0.5SD etc. It is important that the assessed assumptions are clinically plausible and established a-priori from clinical experts in the field. But too often this is done post-hoc and using a single departure from MAR. A variety of departures from MAR should be considered to test sensitivity.

An alternative implementation considers a variety of differences of means (for continuous data) between the observed and missing cases and seeks the point at which the study conclusion is changed. This can be done by sequentially scaling or adding increments to the MAR mean. It is then important to assess the plausibility of the mean which results in changes. This is referred to as a tipping point analysis in Yan (2009) [30].

After specifying and fitting the separate response models for each pattern, they are weighted by their respective probabilities to obtain inference. As discussed, not all response models will be identifiable from the data. By setting inestimable parameters of deviators data distributions to be functions of those observed, the sensitivity of results to departures from MAR can be determined. Either a Maximum Likelihood (ML) or full Bayesian approach can be used to fit pattern mixture models.

White et al. [1] describe how approximate formula for the treatment effect and its standard error can also be employed for inference. For a continuous outcome this entails using a mean score approach, following analysis of the complete data. For example, consider a simple two arm trial setting with a single continuous outcome, denoted Y_{iz} for patient i in arm z where $z = a, r$ (active arm or reference/control arm). One potential pattern mixture model supposes that those observed in the active arm follow a distribution with mean μ_a and variance σ^2 whilst those unobserved in the active arm have mean $\mu_a + \delta_a$ and variance σ_M^2 . For the reference/control arm, those observed come from a distribution with mean μ_r and variance σ^2 . The unobserved in the reference arm have mean $\mu_r + \delta_r$ and variance σ_M^2 . That is,

$$\begin{aligned} Y_{iz}|R_i = 1 &\sim (\mu_z, \sigma^2) \\ Y_{iz}|R_i = 0 &\sim (\mu_z + \delta_z, \sigma_M^2) \end{aligned}$$

Under this model, where $\pi_z = Pr(R_{iz} = 0)$, the average treatment effect is,

$$\Delta = [(1 - \pi_a)\mu_a + \pi_a(\mu_a + \delta_a)] - [(1 - \pi_r)\mu_r + \pi_r(\mu_r + \delta_r)] = (\mu_a - \mu_r) + (\pi_a\delta_a - \pi_r\delta_r).$$

As noted by White et al. [1] $(\mu_a - \mu_r)$ can be estimated as the average treatment effect amongst the completers using the usual complete case analysis. Therefore in this simple setting the δ_z -adjusted treatment effect can be derived in a straight forward manner using the complete case estimate of $(\mu_a - \mu_r)$ and sample estimates of π_a and π_r . White et al. [1] also present formula for establishing the associated variance estimator under the assumption of normality. With additional covariates and longitudinal follow-up the formula are however not so straight forward.

A detailed example which illustrates how to implement a full Bayesian approach to pattern mixture modelling using `winBUGS` is also presented in [1]. In Section 1.5 we will see how MI provides an alternative solution for pattern mixture modelling, which avoids the need to fit the models directly, which can often be quite complex and require sophisticated programming. Essentially the MI analysis approximates a full Bayesian approach. MI can be a more accessible practical approach for busy trialists.

1.3.4 Alternative approaches to sensitivity analysis

An alternative approach to sensitivity analysis entails adding ancillary variables to the analysis of the outcome of interest using MAR methods [31, 12, 32]. That is variables beyond the trials design factors, in order to make the MAR assumption more plausible. This approach seeks to improve the performance of the missing data procedure.

Ancillary variables can be added to likelihood based analyses, included mixed models for repeated measures (MMRM), as covariates or as an additional response. Alternatively, if analysis is conducted via MI, they can be simply added to the imputation model. Taking this approach the ancillary variables are not also required in the analysis model.

A different approach to sensitivity analysis is to assume that the missing values can be represented by a bad outcome. This is most natural for a categorical or binary outcome. For example in an eye injury trial where the outcome is defined as clinical improvement in visual acuity in the study eye (yes/no), sensitivity analysis can explore the impact of the unobserved having no clinical improvement. After completing the missing values the analysis method that would have been used in the absence of missing data can be used.

With a continuous outcome this is not so straight forward. Methods which assign an order to observed and unobserved outcomes have been proposed. The unobserved can be assigned appropriately poor scores to place them in the required order. Statistical methods based on ranks can then be employed [33, 34].

An approach known as Last Observation Carried Forward (LOCF) sets patients missing values equal to their last observed measurement, enabling complete data methods to then be used. Although it is commonly used as a sensitivity analysis [9], we do not advocate this approach (or other single imputation methods). Many researchers, including [7, 35, 36, 37, 38], have highlighted the undesirable consequences of performing such an analysis. In brief, LOCF results in downwardly biased variance estimators, since it is a single imputation method which treats imputed observation as if they were observed. The resulting treatment estimator may be conservative or liberal depending on the context. Further it does not always represent a meaningful or realistic treatment comparison. LOCF has been shown to result in biased treatment estimators under MAR and MCAR [39]. Only under very specific and unrealistic assumptions is LOCF valid [38].

More recently doubly robust methods which utilise IPW have been proposed. Doubly robust estimators require three models: (i) the (analysis) model relating the response to the explanatory variables and covariates of interest, (ii) a model for the probability of response and (iii) a joint model of the mean of partially observed variables given fully observed variables. If either model (ii) or model (iii) are incorrect then doubly robust estimators have the property that the estimators

in model (i) are still consistent. This is because the expectation of the estimating equation is still zero. We do not expand further on this approach here but refer the reader to Chapters 9 and 22 in [8] and [21] for a more in-depth discussion of this approach.

In this thesis we are interested in sensitivity analysis via the pattern-mixture modelling approach (discussed in Subsection 1.3.3), since we believe assumptions are framed in the most accessible form. Additionally, as we outline in following sections, MI provides a direct means for analysis without the need for complex model fitting or formula.

1.3.5 Estimands

Up to now we have focussed discussion on sensitivity analysis, in the general context of assessing the impact of alternative missing data assumptions (i.e. MCAR, MAR or MNAR) on our model based inferences. We now introduce the notion of the estimand of interest. That is the precise definition of what is to be estimated to address the scientific question under study; for who, what and when the trial's intervention effect is estimated for. Therefore, what we wish to explore if inference is sensitive to.

The primary analysis of a trial should be designed to address the estimand of interest. When framing sensitivity analysis in the clinical trial setting we must also carefully consider the precise estimand of interest. This is because the choice made on how to model the missing data, i.e. how to model departures from MAR can influence what is actually being estimated. The missing data model should be consistent with what is being estimated.

Carpenter, Roger and Kenward introduce two main types of estimand it is important to distinguish between [13]. First, a *de-jure* estimand estimates the treatment effect had all patients adhered to the protocol as planned. Thus seeks to answer an efficacy question as to whether the treatment works under the best case scenario. It is closely linked to a per protocol or on-treatment analysis. If we assume missing data are MAR and patients withdraw when they violate the protocol, a likelihood based analysis naturally addresses the *de-jure* question, but is not the only way to do so. Thus the *de-jure* estimand links with MAR. This is in contrast to a *de-facto* estimand which estimates the treatment effect seen in practice, that is the effect of the randomised treatment regardless of protocol adherence. A *de-facto* estimand may be more appropriate where we seek to estimate the holistic effect of intending to treat a group of patients. It is linked to an Intention-To-Treat (ITT) analysis, which includes all patients as randomised, regardless of the actual treatment taken [40].

As discussed by Carpenter, Roger and Kenward these two terms (*de-jure* and *de-facto*) correspond to estimands as defined by the National Research Council [12]. Further they put the focus on the assumption about post-deviation behaviour and provide a common language to talk about missing data and compliance issues in safety and efficacy.

With missing data, in both cases the primary analysis must be consistent with the estimand of interest, in that it must make the most plausible appropriate assumptions about treatment use for cases with missing data following deviation or dropout. Sensitivity analysis to address the impact of alternative missing data assumptions for the primary estimand should then be conducted. In many settings it will be important to explore how robust inferences are to assumptions about

differing post-deviation treatment behaviour in sensitivity analysis. Departures from the MAR assumption can also be formulated in terms of proposed or known treatment group differences. The pattern mixture modelling route lends itself nicely to exploring reference based departures from MAR.

In such cases the underlying missing data assumptions might change the definition of the estimand, to ensure that the robustness of the primary estimand is assessed. This provides a more complete picture of the treatment under investigation. Hence the need to have accessible assumptions to discuss with experts. Often a de-jure estimand will be of primary interest. Clearly defined alternative/secondary estimands, exploring departures from the de-jure/MAR linked analysis provide additional contextually relevant insights. So a generic overall framework for sensitivity analysis in the clinical trial setting, as presented by Kenward in [13] is,

1. A clear definition of the estimand of interest.
2. The assumptions under which the primary analysis is valid for this estimand.
3. A nomenclature for practically relevant and accessible departures from these assumptions.
4. Valid methods for assessing sensitivity to these assumptions.

Where the specific approach taken in step (3) will be guided by the estimands of interest, dictated by the specific setting at hand.

Regardless of whether de-jure or de-facto estimands are of interest, the discussed pattern mixture modelling route provides a useful path for assessing relevant assumptions in step (3). Whilst methods for fitting such MNAR pattern mixture models can often be computationally awkward, MI can be employed to avoid the need of fitting MNAR models directly. We now outline in full the flexible MI procedure, prior to discussing and demonstrating its flexibility for fitting MNAR pattern mixture models.

1.4 Multiple imputation

MI was originally introduced by Rubin in 1987 as a Bayesian procedure with good frequentist properties to handle non-response in sample surveys [19]. The method and its applications to RCTs have since been studied extensively by many including Rubin (1996) [41], Little (1996) [42], Schafer (1997) [43] and Carpenter and Kenward in (2008) [7] and (2013) [44]. The standard MI procedure imputes missing data for a patient as a Bayesian draw from the conditional distribution of their missing data given that observed under the assumption of ignorability (i.e. MAR). This is done in a way that takes full account of the uncertainty created by the missing data, $k = 1, \dots, K$ times to create K completed datasets.

That is, as described by Carpenter and Kenward [44] for $k = 1, \dots, K$ imputations, for patient i data is imputed from the conditional distribution of their missing data given observed,

$$[\mathbf{Y}_{im} | \mathbf{Y}_{io}, \boldsymbol{\eta}], \quad (1.2)$$

where $\boldsymbol{\eta}$ denotes the current draw of the parameters of this distribution from the appropriate Bayesian posterior distribution, unique to the present draw k . For each imputation required we must first obtain $\boldsymbol{\eta}$ prior to drawing \mathbf{Y}_{im} . The posterior of $\boldsymbol{\eta}$ is,

$$[\boldsymbol{\eta} | \mathbf{Y}_o] \propto [\mathbf{Y}_o | \boldsymbol{\eta}] [\boldsymbol{\eta}]. \quad (1.3)$$

For each imputation required we obtain a draw of $\boldsymbol{\eta}$ from (1.3). Then for the missing data,

$$[\mathbf{Y}_m, \boldsymbol{\eta} | \mathbf{Y}_o] = [\mathbf{Y}_m | \mathbf{Y}_o, \boldsymbol{\eta}] [\boldsymbol{\eta} | \mathbf{Y}_o], \quad (1.4)$$

where $[\boldsymbol{\eta} | \mathbf{Y}_o]$ is as specified in (1.3). After drawing \mathbf{Y}_m we can then calculate $\boldsymbol{\theta}(\mathbf{Y}_m, \mathbf{Y}_o)$ which estimates the posterior distribution of our key parameter of interest $\boldsymbol{\theta}$. The analysis model of interest, that would have been used on the full dataset had there been no missing data is used at this stage. A new draw of $\boldsymbol{\eta}$ is obtained from (1.3) and \mathbf{Y}_m from (1.4) for as many imputations required. Thus we end up obtaining K draws of $\boldsymbol{\eta}$, for K draws of missing data to estimate $\boldsymbol{\theta}$, K times. The procedure is appealing since standard complete data methods can be employed post-imputation for inference following the imputation of the missing data. We refer to the analysis model of interest as the substantive model. We make clear here that the substantive model is quite separate to the model used to create the imputations (1.2), which we refer to as the imputation model. MI is a two stage process.

Results across imputed datasets are combined using Rubin's rules [19] to give a single MI estimator for inferential results. For a scalar treatment effect, estimated in each imputed dataset as θ_k with estimated variance σ_k^2 , results are combined as follows,

$$\hat{\theta}_{MI} = \frac{1}{K} \sum_{k=1}^K \hat{\theta}_k, \quad (1.5)$$

with estimated variance,

$$\hat{V}_{MI} = \hat{W} + \left(1 + \frac{1}{K}\right) \hat{B}, \quad (1.6)$$

where,

$$\hat{W} = \frac{1}{K} \sum_{k=1}^K \hat{\sigma}_k^2, \quad (1.7)$$

$$\hat{B} = \frac{1}{K-1} \sum_{k=1}^K (\hat{\theta}_k - \hat{\theta}_{MI})^2. \quad (1.8)$$

The first component of the estimated variance \hat{V}_{MI} (1.6), denoted \hat{W} , averages the estimated variance from each imputed dataset, and is consequently referred to as the within variance. \hat{W} essentially estimates the variance of $\hat{\theta}$ that we see if there is no missing data. Thus Rubin's variance formula depends on the (asymptotic) normality of $\hat{\theta}$ if there is no missing data. The second part, denoted as \hat{B} , quantifies the variance of the imputation estimates over the K imputations and is referred to as the between imputation variance. We see it is the variability across imputed datasets, which is provided by the random components in the imputation model (the imputation model parameter draws, and the subsequent missing data draws unique to each imputation) that is used to adjust the variance estimate upwards. As discussed in Chapter 2 of [44] the term $(1 + \frac{1}{K})$ adjusts for the fact we are conditioning on a finite number of imputations, K .

As detailed extensively by Rubin [19], (1.6) gives us a valid estimate of variance, both for Bayesian and frequentist inference (under conditions we expand upon below). One that is bounded below by the variance of the treatment estimate had the missing data been observed and naturally encapsulates the loss of information. Intuitively it makes sense that by incorporating the between imputation variance this variance estimate reflects the uncertainty due to the missing data, which the single imputation methods described earlier did not.

For $\hat{\theta}_{MI}$ a scalar, to test the hypothesis that $\hat{\theta}_{MI}$ equals some null value, that is $\hat{\theta}_{MI} = \hat{\theta}^0$ we refer the below test statistic,

$$\frac{\hat{\theta}_{MI} - \hat{\theta}^0}{\sqrt{\hat{V}_{MI}}},$$

to a t-distribution with v degrees of freedom, where v comes from a standard Satterthwaite-type approximation [19],

$$v = (K-1) \left[1 + \frac{\hat{W}}{(1 + \frac{1}{K}) \hat{B}} \right]^2.$$

Rubin shows this is more appropriate than referring to the standard normal distribution with a finite number of imputations, since there will be uncertainty in the variance parameter. Confidence intervals for $\hat{\theta}_{MI}$ can then be constructed using the appropriate quantiles of the t-distribution.

As Rubin's rules condition on the number of imputations, estimates, confidence intervals and infer-

ences will be sensible with two or more imputations. However, with a small number of imputations results will be imprecise. Rubin [19] showed that the relative variance i.e. the efficiency of an estimate using only K imputations compared to an infinite number is approximately $(1 + \frac{\lambda}{K})$, where λ is the fraction of missing information. As discussed in [7], 5–10 imputations is sufficient to get a reasonably accurate answer for most applications. For more critical inferences, as in this thesis, 50 or more imputations are recommended [44]. The error in estimating p-values will be smaller with larger numbers of imputations. The between imputation variance will also be more precisely estimated with greater K [6, 45].

1.4.1 Justification for Rubin’s rules

Here, we justify Rubin’s rules from the Bayesian perspective by presenting the details given in Carpenter and Kenward, p. 46–47 [44]. We have observed data \mathbf{Y}_o and missing data \mathbf{Y}_m . We now suppress $\boldsymbol{\eta}$ —the parameters of the distribution of the observed data— interest lies in $\boldsymbol{\theta}$, the key parameters of the substantive model of interest. The missing data and parameters of interest have the following joint posterior distribution,

$$f(\mathbf{Y}_m, \boldsymbol{\theta} | \mathbf{Y}_o).$$

Using standard conditional probability rules this can be partitioned as follows,

$$f(\mathbf{Y}_m, \boldsymbol{\theta} | \mathbf{Y}_o) = f(\mathbf{Y}_m | \mathbf{Y}_o) f(\boldsymbol{\theta} | \mathbf{Y}_m, \mathbf{Y}_o).$$

The missing data can now be separated out, and the marginal posterior for our parameters of interest can be expressed as,

$$f(\boldsymbol{\theta} | \mathbf{Y}_o) = E_{\mathbf{Y}_m | \mathbf{Y}_o} \{ f(\boldsymbol{\theta} | \mathbf{Y}_m, \mathbf{Y}_o) \}.$$

By using the rules of iterated expectations, the posterior mean and variance for $\boldsymbol{\theta}$ are then given by,

$$E(\boldsymbol{\theta} | \mathbf{Y}_o) = E_{\mathbf{Y}_m | \mathbf{Y}_o} \{ E_{\boldsymbol{\theta} | \mathbf{Y}_m, \mathbf{Y}_o} \}, \tag{1.9}$$

$$\text{VAR}(\boldsymbol{\theta} | \mathbf{Y}_o) = E_{\mathbf{Y}_m | \mathbf{Y}_o} \{ \text{VAR}_{\boldsymbol{\theta} | \mathbf{Y}_m, \mathbf{Y}_o} \} + \text{VAR}_{\mathbf{Y}_m | \mathbf{Y}_o} \{ E_{\boldsymbol{\theta} | \mathbf{Y}_m, \mathbf{Y}_o} \}. \tag{1.10}$$

These required quantities can be approximated empirically using imputed data for $\mathbf{Y}_m | \mathbf{Y}_o$. If we

let $\mathbf{Y}_{m,k}$ denote the k^{th} draw of the missing data (for $k = 1, \dots, K$) from the Bayesian predictive distribution $f(\mathbf{Y}_m | \mathbf{Y}_o)$ then (1.9) can be approximated as follows,

$$E(\boldsymbol{\theta} | \mathbf{Y}_o) \approx \frac{1}{K} \sum_{k=1}^K E_{\boldsymbol{\theta}}(\boldsymbol{\theta} | \mathbf{Y}_{m,k}, \mathbf{Y}_o).$$

That is by the average of the estimates of $\boldsymbol{\theta}$ from each imputed dataset. For scalar θ this gives (1.5). Equation (1.10) can be approximated as,

$$\begin{aligned} \text{VAR}(\boldsymbol{\theta} | \mathbf{Y}_o) &\approx \frac{1}{K} \sum_{k=1}^K \text{VAR}_{\boldsymbol{\theta}}(\boldsymbol{\theta} | \mathbf{Y}_{m,k}, \mathbf{Y}_o) \\ &+ \frac{1}{K-1} \sum_{k=1}^K \left(E_{\boldsymbol{\theta}}(\boldsymbol{\theta} | \mathbf{Y}_{m,k}, \mathbf{Y}_o) - \frac{1}{K} \sum_{k=1}^K E_{\boldsymbol{\theta}}(\boldsymbol{\theta} | \mathbf{Y}_{m,k}, \mathbf{Y}_o) \right)^2. \end{aligned}$$

This is the average of the estimates of the variance of $\boldsymbol{\theta}$ from each imputed dataset combined with the variance of $\boldsymbol{\theta}$ across imputed datasets. For large K these quantities are valid approximations for the mean and variance of the posterior distribution. As discussed in [44], p. 50, for small K there is increased uncertainty in the estimated mean of $\boldsymbol{\theta}$. To account for this we must therefore adjust the second term on the right hand side (RHS) —the between imputation variance— by a factor of $\frac{1}{K}$, which for a scalar θ gives us (1.6).

Despite being Bayesian in nature, provided some subtle conditions hold, Rubin’s combination rules also provide valid frequentist inference, in that they provide an estimator which is asymptotically unbiased and an accompanying estimate of variance which can be used to construct confidence intervals with coverage equal to that specified. Rubin outlines the requirements for this as follows, which we elaborate on below.

1. Draw imputations following the Bayesian paradigm as repetitions from a Bayesian posterior distribution of the missing values under the chosen models for non-response and data, or an approximation to this posterior distribution that incorporates appropriate between-imputation variability.
2. Choose models of nonresponse appropriate for the posited response mechanism.
3. Choose models for the data that are appropriate for the complete-data statistics likely to be used - if the model for the data is correct, then the model is appropriate for all complete-data statistics.

[19]

Conditions 1 and 2 essentially refer to what Rubin termed *proper* imputation [19]. They require sampling from a properly defined posterior distribution under a correct model for the non-response and data. That is the imputation models assumptions about the missing data mechanism are

correct, such that our estimate of the treatment effect and its variance, averaged over each imputed dataset, $\hat{\theta}_{MI}$ and \hat{W} are unbiased for the complete-data treatment effect and estimated variance, were these to exist. The estimated variance of our treatment effect across the imputations, \hat{B} , should be approximately unbiased for the sampling variance of $\hat{\theta}_{MI}$ over repeated imputations.

We note that conditions 1–3 also imply that the substantive model is correctly specified for valid frequentist inference.

These conditions incorporate the requirement for *congeniality* between the imputation model and analysis model which, subsequent to the development of MI, was described and explored by Meng [46]. Assume there exists a full Bayesian procedure for obtaining the posterior of θ from the joint data distribution and this is partitioned accordingly and also used to impute the missing data. If the resulting imputation distribution is the same as the predictive distribution obtained by the imputation model then the imputation model and substantive model are said to be congenial. That is, the imputation model is derivable from the joint data distribution. We interpret this as the imputation and analysis model must have the same content and structure and so be formed around the same assumptions to be congenial.

When the substantive model and imputation model do not satisfy this condition, they are described as uncongenial [46]. The validity of Rubin’s variance estimator is not guaranteed when this is the case. Uncongeniality may occur when the imputation model contains more variables or structure than the substantive model. When the imputation model contains a congenial imputation model nested within it, then the imputation model is said to be *richer* than the substantive model. Alternatively the imputation model may lack variables or structure present in the substantive model, then the imputation model is said to be *poorer*.

When the imputation model is richer and the additional information built into the imputation model is correct Meng [46] and Rubin [41] show that the $\hat{\theta}_{MI}$ will be more efficient. Or as termed by Rubin “superefficient” [41]. As the imputer has used additional superior knowledge the sampling variance of the MI estimate will be reduced. Since additional predictors are used in the imputation that are not incorporated into the analysis, Rubin’s variance formula over estimates the sampling variability. Confidence intervals will therefore have greater than nominal coverage. Carpenter and Kenward [47] note that this overestimation is typically not large. Thus practically this tends not to be too much of a disadvantage.

However when additional structure included in the imputation model is not correct, this can have more unwelcome consequences. The imputations will be biased, therefore the results of the analysis will be. Biased estimation in analysis will also occur when the imputation model is poorer than the analysis model. For example if an important predictor of the outcome or an important predictor of missingness is not included in the imputation model. As Schafer clearly outlines [43] the main danger from uncongeniality comes when the imputer makes poorly grounded assumptions, that the analyst does not. It is therefore recommended for the imputation of a variable Y to include all variables that are related to the missingness of Y in the imputation model along with variables related to Y.

1.4.2 Obtaining imputations from a Bayesian posterior

It is the repeated Bayesian draws of the parameters in the imputation model ($\boldsymbol{\eta}$) from (1.3) and the subsequent missing data draws (\mathbf{Y}_m) from (1.4) that allow us to take into account all of the uncertainty when estimating $\boldsymbol{\theta}$. The construction and fitting of an appropriate imputation model (1.2) to obtain these draws (or approximate Bayesian draws) given the observed data is therefore key.

With missingness on a single variable, under MAR, this is fairly straightforward to achieve using regression modelling and uninformative priors. We recall that under MAR (or MCAR) regression models where only the response variable has missingness give valid parameter estimates. To illustrate this we consider a simple example, adapted from [44] p. 41–42, where Y_{i1} and Y_{i2} denote two normally distributed continuous variables collected from a sample of $i = 1, \dots, N$ patients. The target of our analysis, that is the substantive model is the marginal mean of Y_{i2} . However a total of n_d values of Y_{i2} are missing, assumed to be MAR conditional on Y_{i1} ; $N - n_d = n_o$ values of Y_{i2} are observed. In this example, using the above notation, $\boldsymbol{\eta} = \beta_0, \beta_1, \sigma_{2.1}$. To impute the missing values we follow the below procedure,

1. Fit the linear regression model of Y_{i2} on Y_{i1} to the observed data using least squares, $Y_{i2} = \beta_0 + \beta_1 Y_{i1} + e_i$, $e_i \stackrel{iid}{\sim} N(0, \sigma_{2.1})$ to obtain estimates of $\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_{2.1}$.
2. Obtain draws for $\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\sigma}_{2.1}$ from the Bayesian posterior, using uninformative priors by first drawing, $\tilde{\sigma}_{2.1} = \frac{(n_o - 2)\hat{\sigma}_{2.1}}{X}$ where $X \sim \chi_{n_o - 2}^2$, then drawing, $\tilde{\beta}_0, \tilde{\beta}_1$ from,

$$\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \sim N \left\{ \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}; \tilde{\sigma}_{2.1} \begin{bmatrix} n_o & \sum_{i=1}^{n_o} Y_{i1} \\ \sum_{i=1}^{n_o} Y_{i1} & \sum_{i=1}^{n_o} Y_{i1}^2 \end{bmatrix}^{-1} \right\}.$$

3. Missing data can then be obtained by drawing from $Y_{i2,k} = \tilde{\beta}_0 + \tilde{\beta}_1 Y_{i1} + \tilde{e}_{i,k}$, $\tilde{e}_{i,k} \sim N(0, \tilde{\sigma}_{2.1})$.
4. Repeat steps 2 and 3, K times.

We fit our substantive model of interest, the marginal mean of Y_{i2} , which can be derived from the data model (i.e. is congenial) to each imputed dataset and obtain appropriate inference with Rubin's rules (1.5)-(1.6).

With missingness on a single non-continuous variable an alternative logistic, multinomial or ordinal regression model may be used in step 1, depending on the specific type of the variable with missingness. With multivariate data subject to non-response, particularly with a combination of different types of variables, e.g. continuous and nominal, or a non-monotone missingness pattern forming an imputation model is more complex.

Two main routes have been established. In the first a joint multivariate normal (MVN) model is assumed for all variables. Various methods can then be employed to make Bayesian draws from this joint distribution. With a monotone missingness pattern the joint likelihood of the observed data can be readily factorised into independent likelihood functions for each partially observed variable. For example, for three variables Y_{i1}, Y_{i2}, Y_{i3} , with a monotone missingness pattern whereby Y_{i1}

is more observed than Y_{i2} , which is more observed than Y_{i3} their joint distribution, omitting parameterisation, can be expressed as,

$$f(Y_{i1}, Y_{i2}, Y_{i3}) = f(Y_{i3}|Y_{i1}, Y_{i2}) \times f(Y_{i2}|Y_{i1}) \times f(Y_{i1}). \quad (1.11)$$

Under MAR we can validly estimate the parameters of each conditional distribution separately using the observed data. Thus imputing from each of the conditionals in turn gives a valid imputation from the joint distribution. Approximate draws from a Bayesian posterior for the parameters of each conditional, followed by draws of missing data can be made sequentially employing the technique outlined above [22, 44]. At each time point previously imputed variables are included in the imputation model.

If the missingness is not monotone then one or more of the distributions in (1.11) will not be validly separately estimated. A Markov Chain Monte Carlo (MCMC) sampler, such as the Gibbs sampler will be required to simulate random draws from the appropriate posterior [43]. MCMC samplers employ Markov chains to generate draws of random variables from complex distributions.

In our context of interest the Gibbs sampler can be utilised to obtain repeated draws of the parameters in the joint imputation model ($\boldsymbol{\eta}$), and the missing data (\mathbf{Y}). The missing data are essentially regarded as parameters and they are updated in turn. The MCMC sampler is initiated with starting values for the parameters, denoted $\boldsymbol{\eta}^0$, which can be computed as the mean and covariances from the observed data. These are used to obtain starting value for the missing data from the appropriate conditional distribution, denoted \mathbf{Y}^0 . Then in turn the Gibbs sampler proceeds to draw each parameter from (1.3) and the missing data from (1.4), conditional on each other and the observed data.

The sampler is run for a sufficient time, known as the ‘burn-in’, to allow the chain to reach its stationary distribution. The current draw of the missing data at the end of the ‘burn-in’ forms the first imputed dataset with the observed data. The sampler then proceeds with further draws since imputations need to be independent, until another draw of the missing data is taken, together with the observed data to form the second imputed dataset for analysis. The updating of the chain, the ‘burn-between,’ and collection of the imputed data is repeated for as many imputations as required. Carpenter and Kenward [44] recommend a ‘burn-in’ of 1000 iterations with a ‘burn-between’ of 500. The Gibbs sampler is outlined in more detail in [44] p. 306–309 and [43] p. 68–70.

In some circumstances, when the observed data likelihood is difficult to compute, the Expectation Maximisation (EM) algorithm may be used to calculate ML estimates to initiate the sampler [43]. Briefly, the EM algorithm is an iterative algorithm for maximising log-likelihoods, consisting of two steps; the E step and the M step [48]. The E step computes the conditional expectation of the missing data, given the observed data and current estimated parameters, i.e. $E(\mathbf{Y}_{im}|\mathbf{Y}_{io}, \mathbf{X}_i, \boldsymbol{\eta})$. These expectations are substituted for the missing data. The M step then maximises the expected complete data log-likelihood estimation, as if there were no missing data. These two steps are repeated until convergence is achieved. The initial parameter vector for the EM algorithm can be found from a complete case analysis or any simple method of imputation.

As a brief aside, we note that since the EM algorithm is likelihood based it can in fact be used as a

method of inference in itself under the assumption of ignorability for missing data. However unlike MI, it does not yield standard errors of the estimates, additional steps are required as detailed in [22], p. 190–199.

Regardless of whether the missing data pattern is monotone or not, using a joint MVN model for MI does necessitate strong assumptions. If data are skewed, one can consider transformation to approximate normality, then impute, then transform back. Schafer [43] however reports simulations that show imputations drawn under the MVN model are robust to moderate skewness. Additionally Schafer reports the normal model to be a useful tool for imputing ordinal and binary data. Nominal variables can be included in the model with a series of binary dummy variables. Thus, in practice the normal model is useful even when the data are not normal.

Alternatively the second route to obtaining imputations, which has been termed Multiple Imputation by Chained Equations (MICE), or Fully Conditional Specification (FCS), involves a series of univariate conditional models, formed as a regression of each partially observed variable, given all other variables [49]. Each variable can then be imputed in turn in the fashion of a Gibbs sampler. The MCMC method is used to complete the imputation process (drawing parameters and missing data sequentially from their conditional distributions).

Strong assumptions of normality are avoided using this approach. Each univariate imputation model is dependent on the characteristics of the variable being imputed. For example, binary variables can be modelled using a logistic regression and continuous variables modelled using linear regression. It is hoped that the collection of the univariate models converge or correspond to a genuine joint distribution, but there is no guarantee this will be the case. As discussed by Azur et al. [50], MICE lacks the theoretical justification of other imputation approaches. Other research however indicates this is not a principal issue and shows how in practice the method performs well and provides unbiased estimates with appropriate coverage [49, 51]. Recently Zhu and Raghunathan [52] investigated the convergence properties in detail and developed conditions for convergence.

MI is very flexible. Although originally developed in the ignorability (i.e. MAR/MCAR) context, MI can be used in sensitivity analysis to explore departures from MAR, i.e. under MNAR, avoiding the need to fit MNAR models directly. Of course, this does require us to re-evaluate the properties of the usual MI estimators, as we will see in this thesis. In the next section we explore alternative implementations of MI which are particularly relevant for sensitivity analysis of clinical trials.

1.5 Relevant accessible sensitivity analysis via multiple imputation

Rubin first recognised the value of MI for sensitivity analysis in 1987 [19]. The imputation model and substantive model are separate and do not have to be the same. They can reflect different structures, thus departures from a MAR response mechanism can be readily accommodated by altering the imputation model and maintaining the substantive model of interest.

Standard MI, under the assumption of MAR, utilises the conditional distribution of partially observed response data given the observed response data for imputation across all missing data

patterns (1.2). Through altering the values of the parameters of this distribution by missingness pattern we can explore departures from MAR via MI. This approach directly corresponds with the pattern-mixture modelling sensitivity analysis approach outlined in Section 1.3. MI provides an accessible route to explore the impact of such models, without the need for complex model fitting or specification of formula.

For MNAR imputation, for each deviating individual with missing data pattern m_i we require the distribution of their missing outcomes, given their observed data denoted as,

$$[\mathbf{Y}_{mi} | \mathbf{Y}_{oi}, m_i, \boldsymbol{\eta}_m]. \quad (1.12)$$

$\boldsymbol{\eta}_m$ are the parameters of this distribution, whose values differ across missingness patterns, and whose values we first have to estimate, before we can impute missing data from (1.12) using the standard MI procedure. That is for each missing data pattern m , we create K complete datasets by taking a draw of $\boldsymbol{\eta}_m$ from the appropriate Bayesian posterior distribution, $[\boldsymbol{\eta}_m | \mathbf{Y}_o]$. We then draw the missing data from (1.12) using the current draw of $\boldsymbol{\eta}_m$ given \mathbf{Y}_o . Each imputed data set is analysed using the substantive model of interest. Results are then combined using Rubin’s combination rules.

When exploring departures from MAR for each missing data pattern m we choose a form for constructing $\boldsymbol{\eta}_m$ from $\boldsymbol{\eta}$, the parameters of our imputation model under MAR. This is based on a pre-specified rule that reflects a specific assumption. This enables us to assess the impact of alternative missing data assumptions on the trial results in a principled manner, with MAR providing a natural starting point to MNAR exploration.

For longitudinal trials with a number of different missing data patterns, as with all pattern-mixture approaches—and indeed less interpretable selection modelling approaches—this can require the specification of many parameters. Two alternative techniques for forming $\boldsymbol{\eta}_m$ are now discussed for longitudinal clinical trials with continuous outcomes. We will see that explicit parameter specification is however not always required.

1.5.1 The ‘ δ -method’

Quite often it may be of interest to explore the impact of deviators having a poorer response post-deviation than those observed. In other cases it might even be of interest to assess the impact of a better response post-deviation. The ‘ δ -method’ provides a useful accessible route to explore such pertinent departures from MAR via MI in the clinical trial setting.

For each missing data pattern m , the parameters of the conditional data distribution used for imputation, $\boldsymbol{\eta}_m$, are constructed using the parameters of the MAR implied conditional distribution, $\boldsymbol{\eta}$, and numerical information which ideally would be elicited from experts. That is a consensus is reached on the extent to which the parameters of the distribution of missing data are likely to differ from the observed. The postulated difference for pattern m is referred to as δ . $\boldsymbol{\eta}$ is then edited accordingly (by δ) to obtain $\boldsymbol{\eta}_m$.

Carpenter and Kenward [7, 44] outline how this pattern mixture MI approach can be used to accommodate an anticipated change in rate of improvement or decline post-deviation from that predicted under MAR in a longitudinal trial setting. They introduce a chronic asthma randomised placebo controlled trial. The primary outcome was FEV₁ (Forced Expiratory Volume in 1 second, measured in Litres) recorded at baseline, 2, 4, 8 and 12 weeks: however a number of patients failed to complete the study in both arms. Further those who dropped out were observed to have a poorer response over times they were observed. Initially K imputations are generated under MAR. For each deviating patient, the first MAR imputed FEV₁ observation is reduced by a postulated amount, δ , which itself has an appropriate prior distribution. The second MAR imputed FEV₁ observation is reduced by twice this postulated amount, 2δ , and so on. This reflects a worsening response over time.

Figure 1.1 illustrates the ‘ δ -method’ using data from a similar asthma trial with FEV₁ outcome data (described in full detail in Section 1.7.2). The first MAR imputed FEV₁ observation is reduced by $\delta = 0.05$, and the second by $2\delta = 0.10$.

The δ -adjusted imputed datasets are analysed using the substantive model of interest. Results are combined across imputed datasets using Rubin’s rules. In the present example this provides the treatment estimate where active deviators are assumed to have a change of $\delta = 0.05$ in the rate of decline post-deviation, i.e. a gradual worsening in response over time. Full technical details of the ‘ δ -method’ are presented in Chapter 5.

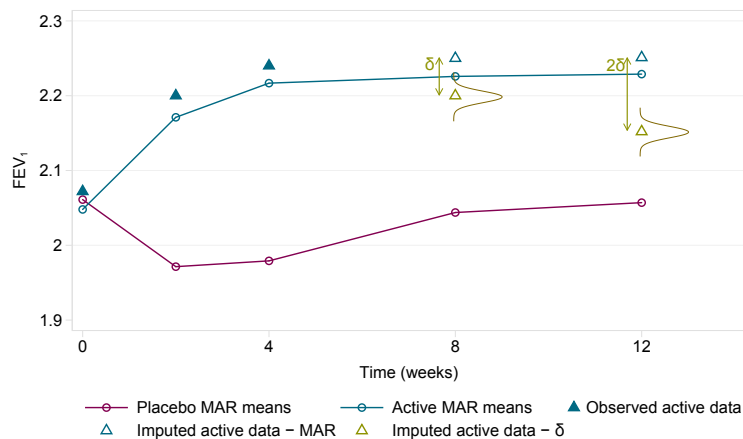


Figure 1.1: Illustration of the ‘ δ -method’

Such formal pre-specification of quantitative sensitivity parameters that is required in the regulated trial environment is however notoriously difficult [28]. Even more so when additional baseline covariates are included in the analysis. A different approach can be employed, whereby adjustments for δ of an increasing size can be applied until the conclusions of the trial change. Afterward the plausibility of the assumption underlying the analysis when conclusions change can be evaluated, that is the particular size of δ representing the difference in outcome between the observed and non-observed. This is the tipping point analysis approach [30]. However this approach is open to criticism since it often relies on post-hoc justification.

1.5.2 Reference based sensitivity analysis

Another option is to make statements about post-deviation data, by reference to other groups of individuals in the trial (typically individuals in different treatment arms). For each missing data pattern we can construct the parameters for our missing data, η_m , using within trial information. The ‘ δ -method’ can require pre-specification of a large number of sensitivity parameters especially when the data are longitudinal and the number of missing data patterns (m) is large. Reference based approaches avoid the need for explicit parameter specification, which can often be hard to justify. In-study data is used to make qualitative rather than quantitative missing data assumptions based on plausible clinical scenarios.

An early version of reference based MI was originally introduced by Little and Yau in 1996 [42]. Little and Yau used sequential regression and MI to impute for patients as actually treated after dropout. Building on the ideas of Little and Yau, more recently in 2013, Carpenter, Roger and Kenward [13] formalised the approach and presented a novel collection of MI procedures for reference based sensitivity analysis of trials with protocol deviation.

The technique revolves around a fitted MAR model to the observed data for each treatment arm. The parameters of the required conditional distributions used to impute data, η_m , are pieced together with qualitative reference to trial arms, using the MAR parameters. For continuous data the mean vector and variance-covariance matrix of the required conditional distributions are constructed using the information from other groups in the trial. The primary analysis model, based on a comparison of the randomised groups, is retained in the sensitivity analysis, in keeping with the ITT principle. This allows for the assessment of the impact of alternative sampling behaviour only on the primary analysis as originally planned.

For example, future statistical behaviour of deviating patients on an active treatment can be assumed to be similar to that of observed control subjects. In the chronic asthma trial comparing an active treatment against placebo, it may be of interest to assess the robustness of inferences from the primary analysis if we assume deviating patients in the active arm stopped taking their treatment post-deviation i.e. the active patients jump to placebo behaviour. Under this de-facto assumption the unobserved data for deviating active patients can be imputed from the appropriate conditional distribution formed using the mean and covariance matrix from the active arm at pre-deviation times and the mean and covariance matrix from the placebo arm for post-deviation times. Figure 1.2 is a schematic illustration of so-called jump to reference (J2R) imputation for the same active deviator observed in Figure 1.1. In comparison to imputation under MAR (Figure 1.1), we see the imputed data is much lower under the J2R assumption.

A full description of the reference based MI procedures, with technical details, is given in Chapter 2.

The above example illustrates how MI provides an accessible route for open and interpretable sensitivity analysis that all trial personnel can understand. No direct estimation of a MNAR model is required. Both de-jure and de-facto estimands can be assessed via this route. It is a useful tool to assess both the sensitivity of the primary estimand to alternative missing data assumptions and the sensitivity of the primary estimand to alternative target estimands.

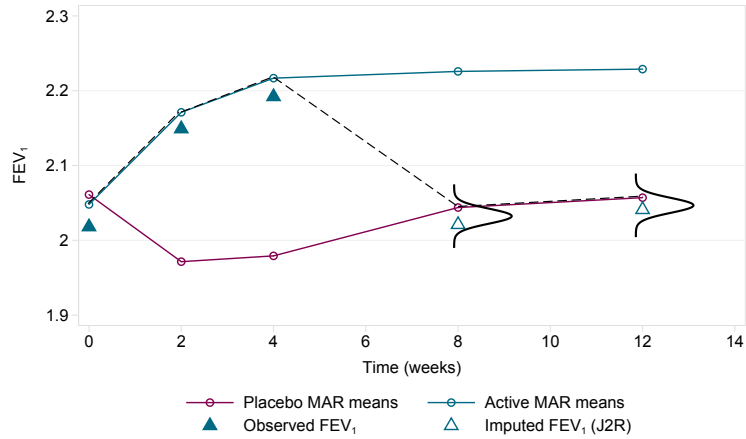


Figure 1.2: Illustration of jump to reference imputation

1.6 Two classes of sensitivity analysis

We have described sensitivity analysis using the definition provided by Daniels and Hogan as an “assessment of sensitivity of model-based inferences to assumptions that cannot be verified or checked within the data” [4]. For the purpose of the following chapters it is important to make clear a final distinction between two different types of sensitivity analysis that fall under this working definition when we are assessing the impact of alternative missing data assumptions [13].

First where the original primary analysis is retained, despite it being *inconsistent* with the postulated data mechanism. For example reference based MI, retains analysis as randomised, despite alternative assumptions about treatment taken being made in the imputation model. This type of sensitivity analysis seeks to assess exclusively the impact of alternative assumptions about the missing data on the original intended analysis. This is a practical approach which is desirable since it shows how inference from the primary analysis varies over different assumptions.

Second where the analysis model is chosen to satisfy the postulated data mechanism in the particular sensitivity analysis. In this case the entire analysis is built upon the postulated MNAR assumption. Complex MNAR modelling will be required. This type of sensitivity analysis seeks to address the impact of reconstructing the entire analysis under alternative assumptions.

MI is useful for both classes of sensitivity analysis. For the later type the imputation model and analysis model are congenial. For the former category, which includes the reference based procedures, the imputation model and analysis model are not necessarily congenial. There will be no full likelihood analysis that corresponds to this analysis. This distinction is important as described by Kenward in [16] “Failure to appreciate this difference can lead to mistaken attempts to assess the properties of one approach through the behaviour of the other.” We are interested in evaluating the first category of sensitivity analysis in this thesis where the imputation model and substantive uncongenial. Because whilst the imputation model and substantive analysis model are separate and can be different, as alluded to in Section 1.4 Rubin’s rules are not necessarily guaranteed to hold when the imputation and substantive model have different structures, i.e. they are uncongenial. We now introduce two case studies that serve as motivation and illustration for the theory developed in this thesis.

1.7 Motivating datasets

1.7.1 Peer review study

The quality of peer review is a controversial topic. Schroter et al. [53] performed a single blind randomised controlled trial among reviewers for the BMJ (British Medical Journal) to investigate the effects of face-to-face training or a self-taught training package on the quality of peer review. As written in the original trial publication [53] and by others [44], each recruited reviewer was sent a baseline article to review (paper 1). Upon return of the review as randomised reviewers then (i) participated in a full day of face-to-face training or (ii) received a self-taught training package in the post or (iii) received no training. Two to three months later participants who completed their first review were sent a further article to review (paper 2).

Participants were asked to review articles based on previously published articles with nine major and five minor errors introduced. Original author names, titles and location were changed. The quality of review was measured by the mean (from two raters) of the validated Review Quality Instrument (RQI). The RQI contains eight items, each scored from one to five. Analysis of the trial indicated that the only significant difference was in the quality of the review of paper 2 between those in the self-taught group and no training group. In this thesis we therefore focus our attention on the self-taught group versus no training group comparison for paper 2. The response in our analysis is the mean of the first seven items of the RQI, averaged over the two raters. This ranged between 1 and 5, where 5 indicates a perfect score.

Table 1.1 summarises the observed review quality data at baseline for those who completed the second review and those who did not. We can see that a larger proportion of participants in the self-taught group failed to return paper 2, further these participants had a notable lower baseline review quality score, i.e. a disproportionate number of poor reviewers in the self-taught group failed to review paper 2.

	No training			Self-taught		
	n	mean	SD	n	mean	SD
Returned paper 2	162	2.65	0.81	120	2.80	0.62
Did not return paper 2	11	3.02	0.50	46	2.55	0.75

Table 1.1: Peer review study: quality of peer review at baseline for those who did and did not complete the second review

The original trial analysis was conducted under the MAR assumption. It is important to assess how robustness the conclusions of the primary analysis are to the MAR assumption. Using reference based MI we are able to establish what the results would like if we assume the reviewers who did return paper 2 behaved like those in the no training group. In Chapter 6 we will apply the reference based methodology evaluated in this thesis to establish the impact this assumption has on inference from the primary analysis. We will compare and contrast this result with sensitivity analysis via the ‘ δ -method,’ which has previously been conducted by Carpenter and Kenward [44]. We replicate and extend their δ -adjusted analysis.

1.7.2 Asthma trial

Data from a randomised double-blind placebo controlled trial of budesonide delivered by Turbuhaler for the treatment of adult patients with chronic asthma [54] provides further motivation for the evaluation of the reference based MI sensitivity analysis procedures.

As described in [13, 54, 55], a total of 473 individuals were randomised to either a daily dose of 200, 400, 800 or 1600 μg of budesonide or placebo. The primary outcome—measured at weeks 0 (baseline), 2, 4, 8 and 12—was forced expiratory volume in one second (FEV_1), recorded in Litres (L), however a number of individuals deviated and did not complete the full 12-week follow-up.

In this thesis we focus our attention on the placebo and lowest dose active arms (200 μg budesonide). The observed mean profiles by treatment arm and the various missing data patterns are shown in Figure 1.3. Only 38 of the 92 individuals in the placebo arm (41%) and 72 of the 91 individuals in the active arm (79%) remained in the trial at 12 weeks; 3 individuals (2 placebo and 1 active) had interim missing data.

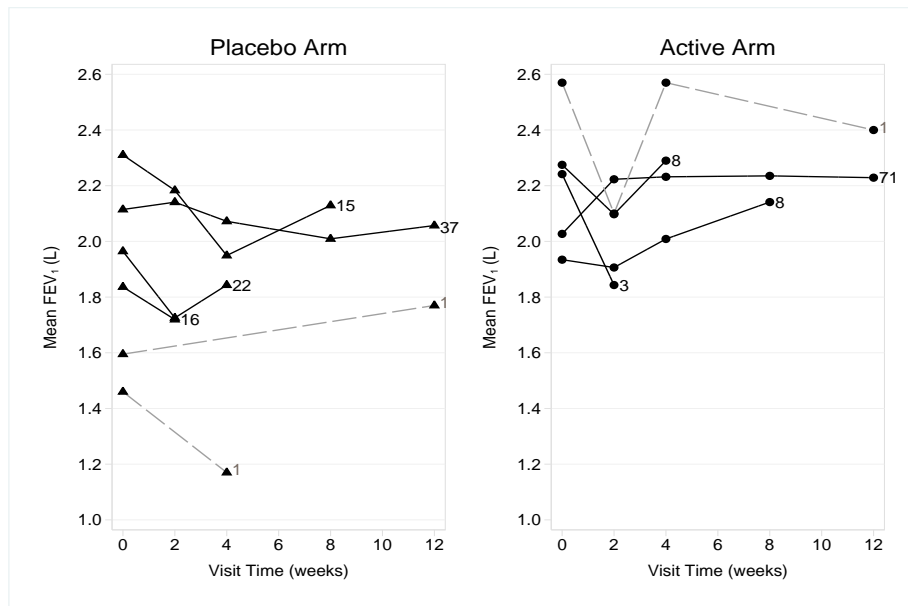


Figure 1.3: Observed Mean FEV_1 by treatment arm and deviation profile against time. Solid lines join observed means at each time point for the various deviation (withdrawal) patterns; dashed lines join observed means of the three individuals with interim missing data. Numbers indicate the counts of individuals with the associated profile.

The primary analysis of the original trial consisted of a linear regression of the 12 week FEV_1 outcome on-treatment group, adjusted for baseline FEV_1 , using data from the 110 individuals measured at week 12. This gave a treatment effect of 0.239 L, $p = 0.017$. Carpenter and Kenward [44] undertook an improved analysis under MAR including all the observed data which gave a treatment effect of 0.335 L, $p = 0.002$. Reference based sensitivity analysis allows us to assess the robustness of the results to various plausible post-deviation assumptions.

In Chapter 6 we illustrate the relevance and accessible nature of the evaluated reference based sensitivity analysis methodology with the newly developed Stata command “mimix” [56]. The asthma trial dataset also informs the simulation studies in Chapters 2 to 5. In the remainder of this chapter we outline the focus of the evaluations in this thesis and the structure of the following chapters.

1.8 Thesis focus

The main focus of this PhD is the statistical performance of the reference based MI procedures introduced by Carpenter, Roger and Kenward [13] for sensitivity analysis of longitudinal RCTs with continuous outcome data. Specifically, the evaluation of Rubin's variance estimator is of key interest. In the reference based sensitivity analysis settings the imputation model and substantive analysis model make different assumptions. Rubin's rules are not necessarily guaranteed to hold when this is the case.

As outlined in Subsection 1.5.2 reference based MI is a promising technique for relevant accessible methods of sensitivity analysis. As highlighted in [11] and [12] there is a need for such methods. We confirm that the reference based MI estimator is unbiased and define a criteria for the variance of reference based (MI) estimators. We show how Rubin's variance estimator relates to this, to fully establish the methodology.

In Chapter 2 we provide technical details on the reference based MI procedure for a continuous outcome, along with elaborated reasoning for evaluating the variance estimate in this setting. This includes a motivating simulation study. Chapter 3 introduces an information anchoring principle for variance estimation in sensitivity analysis of the type proposed. We then investigate the usual MI variance estimate in the baseline and single follow-up trial setting. This leads to a general proposition and proof on the properties of the usual MI variance estimate in this context and more complex longitudinal trial settings in Chapter 4. We then evaluate the usual MI variance estimate for the ' δ -method' in Chapter 5 for comparison. Software for the proposed method of reference based sensitivity analysis is introduced in Chapter 6 and applied to the two case studies. This chapter is based on a Stata journal article, authored by myself and my supervisors [55]. We finish with a summary of the results in Chapter 7 and discuss other proposals.

Chapter 2

Reference based sensitivity analysis via multiple imputation

In Chapter 1 relevant accessible approaches to sensitivity analysis for trials with missing data via MI were introduced. Through modifying the MAR imputation distribution the impact of alternative missing data assumptions on the primary analysis model can be assessed. Following Carpenter, Roger and Kenward [13] we argued that one principled way of doing this, particularly relevant for the RCT setting, is through reference to other trial arms.

In this chapter we first outline in full detail the novel collection of MI procedures proposed by Carpenter, Roger and Kenward [13] in Section 2.1, so there is a clear foundation for the theory that follows. These procedures construct post-deviation data distributions for imputation with qualitative reference to relevant trial arms, for referenced based sensitivity analysis of a continuous outcome. We then lay the motivation for exploring the properties of these procedures, with specific focus on the variance estimate in this setting in Section 2.2. A general principle for variance estimation in this setting is proposed. In Section 2.3 the results of an initial investigatory simulation study are presented and discussed.

2.1 Reference based multiple imputation

Carpenter, Roger and Kenward [13] present a novel collection of MI procedures that can be utilised for performing reference based sensitivity analysis of trials with missing data. As outlined in [13] the primary analysis is conducted under the MAR assumption, where the conditional distributions of later response data given earlier response data are equal for patients who do and do not deviate. Sensitivity is explored by retaining the primary analysis model and fitting this to data imputed under a range of clinically relevant assumptions.

The sensitivity analysis builds out from the fitted MAR model, and the parameters of the required conditional distributions used to impute MNAR data are pieced together with qualitative reference to trial arms. For continuous data the mean vector and variance-covariance matrix of

the required conditional distributions are constructed with reference to other arms. The pieced together conditional distributions are then used to multiply impute K completed datasets.

The generic reference based MI algorithm of Carpenter, Roger and Kenward, for longitudinal trials with a continuous outcome, as described in [13] is now presented with minor modifications.

1. Separately for each treatment arm take all the observed data, and assuming MAR, fit a MVN distribution with an unstructured mean (i.e. a separate mean for each of the baseline and post-randomisation observation times) and variance-covariance matrix using a Bayesian approach with an improper prior for the mean and an uninformative Jeffrey's prior for the covariance matrix.
2. Draw a mean vector and covariance matrix from the posterior distribution for each treatment arm. Specifically we use the Markov-Chain Monte Carlo (MCMC) method to draw from the appropriate Bayesian posterior, with a sufficient burn-in and update the chain sufficiently in-between to ensure subsequent draws are independent, given the observed data. The sampler can be initiated using the EM algorithm.
3. Use the draws in step 2 to form the joint distribution of each deviating individual's observed and missing outcome data as required. This can be done under a range of assumptions, in order to explore the robustness of inference about the treatment effects. The options presented by Carpenter, Roger and Kenward in [13] that each translate to a relevant assumption are described in Table 2.1.
4. Construct the conditional distribution of missing (post-deviation) given observed outcome data for each individual who deviated, using their joint distribution formed in step 3. Sample missing post-deviation data from the conditional distributions to create a completed dataset.
5. Repeat steps 2–4 K times, resulting in K imputed datasets.

Example

To illustrate how the generic algorithm enables reference based sensitivity analysis we consider the chronic asthma RCT introduced in Section 1.7.2. This was a double-blind RCT comparing the active treatment budesonide against placebo. The primary outcome was FEV₁ recorded at baseline (time 0), 2, 4, 8 and 12 weeks. Using the approach described above, initially we fit a MVN distribution, assuming MAR, with an unstructured mean and variance-covariance matrix separately by treatment arm. We then obtain a draw of the MAR mean vector and variance-covariance matrix by treatment arm via the MCMC method from the fitted Bayesian posterior. We denote the current draw of the active group means and variance-covariance by $\boldsymbol{\mu}_a = [\mu_{a0}, \mu_{a2}, \mu_{a4}, \mu_{a8}, \mu_{a12}]$ and $\boldsymbol{\Sigma}_a$. The current draw of the placebo group means and variance-covariance from the posterior is denoted by $\boldsymbol{\mu}_p = [\mu_{p0}, \mu_{p2}, \mu_{p4}, \mu_{p8}, \mu_{p12}]$ and $\boldsymbol{\Sigma}_p$. Figure 2.1 illustrates the current draw of the mean MAR vector by treatment arm for the asthma data. The three solid triangles in Figure 2.1 represent observations from a randomly selected active patient who deviated some time following week 4 and henceforth has unobserved data.

Under the primary MAR assumption the joint distribution of a deviating patient's observed and

missing data is formed as MVN with mean and variance-covariance matrix from their randomised arm. For the randomly selected active deviator in our example, their joint distribution is formed as MVN with mean μ_a and covariance matrix Σ_a as shown in Figure 2.2. For each deviating patient we then construct the conditional distribution the joint MAR distribution implies for their missing data, given their observed data. A random sample is drawn from each of the formed conditional distributions to complete the deviating patient's measurements for imputation K . The process of drawing a mean vector and covariance matrix (for each treatment arm) and forming the required conditional distribution for subsequent imputation is repeated for as many imputations required. The primary analysis model which would have been used in the absence of any missing data is fitted to each imputed data set and results combined using Rubin's rules for inference.

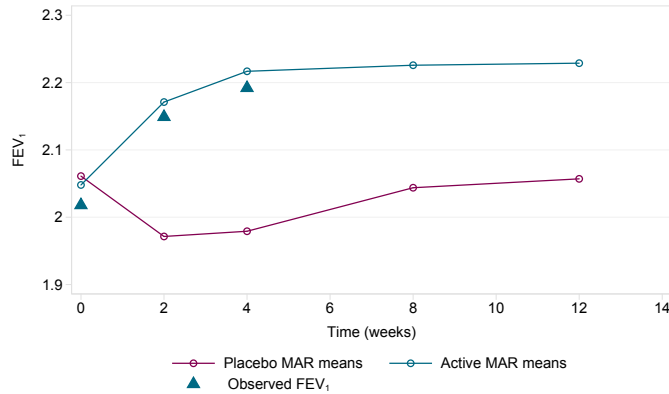


Figure 2.1: MAR mean profiles in the asthma RCT

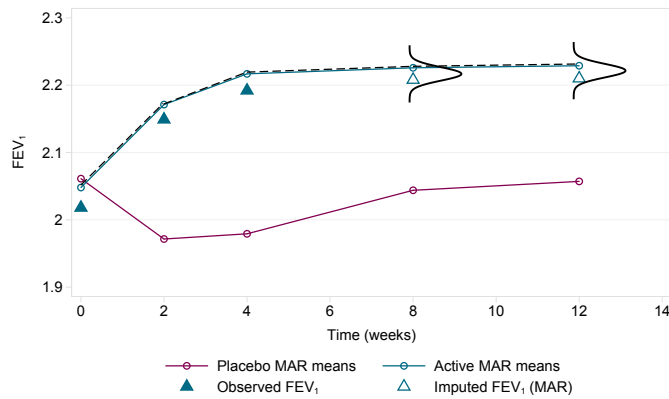


Figure 2.2: Drawing imputed data under MAR in the asthma RCT

In sensitivity analysis we could alternatively assume deviating patients stopped taking their randomised treatment following deviation and subsequently behaved like a member of an alternative reference group. This assumption is referred to as jump to reference (J2R). The joint distribution of a deviating patient's observed and missing data is formed as MVN with mean from the patient's randomised arm for pre-deviation measurements. For post-deviation measurements the mean matches that observed for a reference group. The variance-covariance matrix matches that observed for the randomised arm for pre-deviation measurements and the reference group for the post-deviation and conditional components of post- given pre-deviation measurements. For any deviators already in the reference group this means their missing data will be imputed under MAR. The proposed mean for the deviating active patient in our example, denoted by μ_i , where the

reference group is the placebo arm, is $\boldsymbol{\mu}_i = [\mu_{a0}, \mu_{a2}, \mu_{a4}, \mu_{p8}, \mu_{p12}]$ as shown in Figure 1.2.

The proposed variance-covariance matrix,

$$\boldsymbol{\Sigma}_{J2R} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

as shown by Carpenter, Roger and Kenward [13], is constructed by first partitioning the posterior draws of the active and placebo variance-covariance matrices by their pre- and post-deviation measurements. Below $\boldsymbol{\Sigma}_a$ and $\boldsymbol{\Sigma}_p$ have been accordingly partitioned; 1 indexes pre-deviation measurements (baseline, week 2 and week 4) and 2 indexed post-deviation measurements (week 8 and 12).

$$\boldsymbol{\Sigma}_a = \begin{bmatrix} \mathbf{a}_{11} & \mathbf{a}_{12} \\ \mathbf{a}_{21} & \mathbf{a}_{22} \end{bmatrix},$$

$$\boldsymbol{\Sigma}_p = \begin{bmatrix} \mathbf{p}_{11} & \mathbf{p}_{12} \\ \mathbf{p}_{21} & \mathbf{p}_{22} \end{bmatrix}.$$

Then as shown by Carpenter, Roger and Kenward in [13], $\boldsymbol{\Sigma}_{11} = \mathbf{a}_{11}$, $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21} = \mathbf{p}_{21}\mathbf{p}_{11}^{-1}\mathbf{a}_{11}$ and $\boldsymbol{\Sigma}_{22} = \mathbf{p}_{22} - \mathbf{p}_{21}\mathbf{p}_{11}^{-1}(\mathbf{p}_{11} - \mathbf{a}_{11})\mathbf{p}_{11}^{-1}\mathbf{p}_{12}$.

It may be more appropriate to assume deviating patients maintain a benefit from their randomised treatment and post-deviation their mean profile tracks the mean profile in a reference group. This assumption is referred to as copy increments in reference (CIR). The joint distribution of a deviating patient's observed and missing data is formed as MVN with mean from the patient's randomised active arm for pre-deviation measurements. For post deviation measurements the mean follows the increments observed for a specified reference group. The variance-covariance matrix matches that observed for the randomised arm for pre-deviation measurements and the reference group for the post-deviation and conditional components of post- given pre-deviation measurements. That is $\boldsymbol{\Sigma}_{CIR} = \boldsymbol{\Sigma}_{J2R}$. For any deviators already in the reference group this means, like J2R, their missing data will be imputed under MAR. For our selected active deviator under CIR, where the reference group is the placebo arm, $\boldsymbol{\mu}_i = [\mu_{a0}, \mu_{a2}, \mu_{a4}, \mu_{a4} + (\mu_{p8} - \mu_{p4}), \mu_{a4} + (\mu_{p12} - \mu_{p4})]$, as illustrated in Figure 2.3.

Or we could assume deviating patients never started taking their randomised treatment and hence behaved like a patient from a specified reference group for the duration of the trial. This is the copy reference (CR) assumption. The joint distribution of a deviating patient's observed and missing data is formed as MVN with mean and variance-covariance from a specified reference group for all measurements. Deviators already in the reference group will again be imputed under MAR. For our deviating active patient, as shown in Figure 2.4, $\boldsymbol{\mu}_i = [\mu_{p0}, \mu_{p2}, \mu_{p4}, \mu_{p8}, \mu_{p12}]$ under CR where the reference group is the placebo arm. The variance-covariance matrix is, $\boldsymbol{\Sigma}_{CR} = \boldsymbol{\Sigma}_p$.

Finally we could assume that post-deviation, deviating patients response's stay, on average, constant at the value of the mean for their randomised arm at their last pre-deviation measurement. The joint distribution of a deviating patient's observed and missing data is formed as MVN with mean from the patients randomised active arm for pre-deviation measurements. For post-deviation measurements the mean remains constant, at the level of the patients randomised arm, for their last pre-deviation measurement. The variance-covariance matrix matches their randomised arm. Figure 2.5 illustrates this assumption, which is referred to as last mean carried forward (LMCF), for our active deviator. That is $\boldsymbol{\mu}_i = [\mu_{a0}, \mu_{a2}, \mu_{a4}, \mu_{a4}, \mu_{a4}]$. The variance-covariance matrix matches the active arm, i.e. $\boldsymbol{\Sigma}_{LMCF} = \boldsymbol{\Sigma}_a$.

Under all the alternative reference based assumptions, after forming the required joint distribution for each deviator we construct the appropriate conditional distributions the pieced together MVN distributions imply for the missing data, given the observed data. A random sample is drawn from the formed conditional distributions to complete the deviating patients' measurements for imputation K . Subsequently a new draw of the MAR mean vector and variance-covariance matrix by treatment arm via the MCMC method from the fitted Bayesian posterior is made. The formation of the required distributions and drawing of data is repeated, for the number of imputations required.

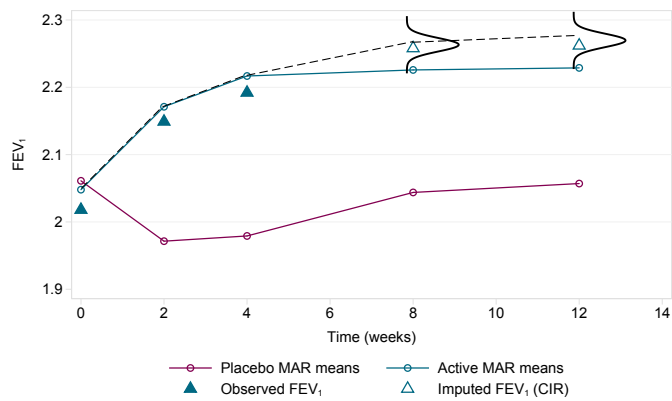


Figure 2.3: Drawing imputed data under CIR in the asthma RCT

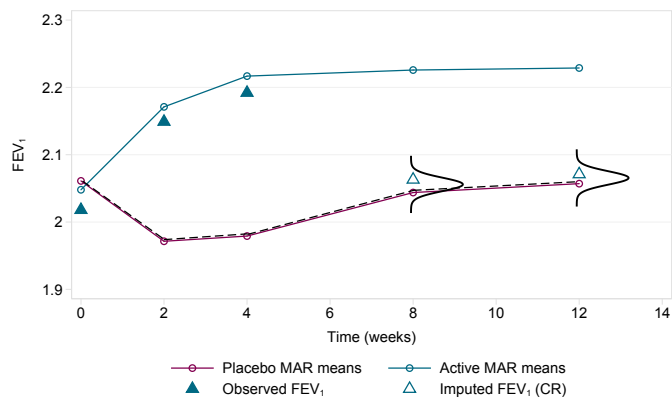


Figure 2.4: Drawing imputed data under CR in the asthma RCT

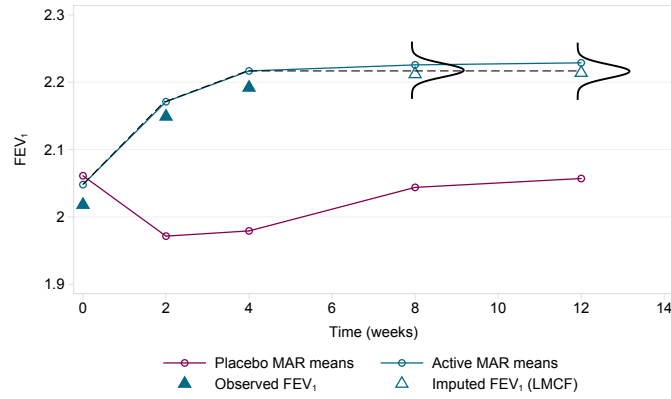


Figure 2.5: Drawing imputed data under LMCF in the asthma RCT

The substantive analysis model using in the primary analysis, which defines treatment group by randomisation, is retained in the sensitivity analysis and fitted to each imputed dataset in turn. This is regardless of the treatment regime assumed in imputation, to assess the impact of alternative sampling behaviour on the primary analysis as originally planned. This follows the Intention-To-Treat (ITT) principle which states that all randomised patients in a clinical trial should be included in the analysis according to the group they are randomised, irrespective of treatment taken [40]. This preserves the randomisation, which ensures treatment groups are alike in all important aspects and only differ with regards to treatment (any differences will be chance imbalances). Carpenter et al. [57] term such an analysis a *design based analysis* and note this also follows the general principle that analysis should not condition on post-baseline information.

The results are then summarised for inference using Rubin’s rules. For example, often the design based analysis model is an analysis of covariance in which the final outcome is regressed on randomised group, adjusted for baseline. Estimates of the treatment effect, $\hat{\theta}_k$ are obtained with variance $\hat{\sigma}_k^2$. Results across imputations can then be combined using Rubin’s rules (1.5) to (1.6) to estimate the overall treatment effect and its associated variance under the given assumption.

It is important to note that as the design based analysis used in the primary analysis is retained in the sensitivity analysis, the analysis model is not always consistent with the data generation model. Under MAR, post-deviation data for deviating patients will be assumed to behave like that of their original randomisation. Under CR, J2R and CIR post-deviation data is assumed to follow that of an alternative treatment group. Under LMCF post-deviation data are assumed to follow the distribution of earlier outcomes in the patient’s randomised arm. Table 2.1 summarises the options presented by Carpenter, Roger and Kenward [13] for piecing together post-deviation data distributions.

Method	Description
Randomised-arm MAR	The joint distribution of patients' pre- and post-deviation outcome data is MVN with mean and covariance matrix from their randomised arm. A natural option for a de-jure estimand.
Jump to reference (J2R)	The joint distribution is MVN with mean vector from the patients randomised arm up to their last observation time, post-deviation the mean vector follows that observed for a reference group (typically control). The covariance matches the randomised arm for pre-deviation measurements and the reference arm for the conditional components of post- given pre-deviation measurements. Appropriate when we believe the deviator ceased their randomised treatment and started treatment similar to that available in one of the other arms (the reference) post-deviation.
Last mean carried forward (LMCF)	The joint distribution is MVN with mean vector from the patients randomised arm up to their last observation time, post-deviation the means are set equal to the marginal mean for the patients randomised arm at their last observed time. The covariance matrix remains as that for their randomised treatment arm. Appropriate when we believe the effect of randomised treatment is maintained on average post-deviation.
Copy increments in reference (CIR)	The joint distribution is MVN with mean vector from the patients randomised arm up to their last observed time, post-deviation the patients mean increments follow those from a reference arm. The covariance is the same as in J2R. Appropriate when we wish to assume that post-deviation the disease resumes the course observed in the reference arm.
Copy reference (CR)	The joint distribution of patients' pre- and post-deviation outcome data is MVN with mean and covariance matrix from a reference arm regardless of deviation time. A natural option when we believe patients followed a different (reference) treatment from their randomised allocation throughout the trial.

Table 2.1: Reference based multiple imputation options

2.2 Variance estimation

2.2.1 Motivation for evaluation

It is important to assess the statistical properties of the reference based multiple imputation procedures to ensure their use is fully justified and theoretically sound.

Under the discussed reference based MI options patients are assumed to follow an alternative treatment regime to their original randomisation and data is generated under that alternative regime. The design based analysis model however does not incorporate these assumptions and treats all patients as originally randomised, regardless of deviation. It is important that the design based analysis model used in the primary analysis is retained in the sensitivity analysis. Because the impact of alternative assumptions about the missing data on the key parameter of interest in the primary analysis can then be assessed.

In reference based MI, Rubin's rules are used to combine results across imputed datasets, but where there is a mismatch between the data generation and analysis model, it is actually unclear precisely what Rubin's variance formula is estimating. As outlined in Section 1.4.1 Rubin's rules were developed in a congenial setting. That is, the imputation and analysis model are both derivable from the joint distribution of the data and missingness. They consequently make the same assumptions and have the same structure.

But here we are in an uncongenial setting. There is a mismatch between the data generation and analysis model. They essentially match on the observed data, but the imputation model has structure that is additional to the substantive model for the unobserved data. For example under the J2R assumption, alternative treatment group behaviour is assumed post-deviation for deviators. It is therefore unclear what Rubin's variance estimator will give us in the reference based settings.

It is possible we could have more information in the sensitivity analysis than we would have had if we had actually observed the data. On the other extreme, it is quite possible to be so imprecise there is much less information in the sensitivity analysis than in the primary analysis (conducted under MAR), and any conclusion from the primary analysis can be overturned. Pharmaceutical companies rightly want to be sure that sensitivity analyses are not throwing away expensively obtained information. Regulators wish to be sure that sensitivity analyses are not injecting information unawares. We aim to establish whether Rubin's treatment estimator and variance estimator provide us with appropriate inference in the reference based settings.

In order to answer this question we first need to consider what is required of our variance estimate in the reference based sensitivity analysis setting. Then we can evaluate the proposed use of Rubin's rules.

2.2.2 A lower bound for variance estimation

We now introduce a simple extended example to set out the issues raised and explore our principles for variance estimation. Consider a two arm randomised trial of an active versus reference treatment, denoted by $z = a, r$, consisting of two patients per arm ($i = 1, 2$). Each patient has a single follow-up, Y_{zi} . We assume independence between patients and the data come from normal distributions with a different mean but the same variance in each arm, $Y_{zi} \stackrel{iid}{\sim} N(\mu_z, \sigma^2)$. The design based analysis is the difference in means between the two randomised arms, with variance estimated for each arm using the usual sample variance formula, which is consistent with the underlying data generating mechanism. We denote the estimated treatment effect as $\hat{\theta}$, the estimated variance of the treatment effect by V and the long-run sampling variance by $\text{VAR}[\hat{\theta}]$.

Suppose all patients are observed without deviation, i.e. on-treatment. We are interested in the mean treatment group difference, $\hat{\theta} = (Y_{a1} + Y_{a2})/2 - (Y_{r1} + Y_{r2})/2$. Under our data generating mechanism $\hat{\theta}$ has expectation, $E[\hat{\theta}] = \mu_a - \mu_r$.

The empirical long-run sampling variance of the treatment effect under our assumptions for the data is,

$$\text{VAR}[\hat{\theta}] = \frac{\sigma^2}{2} + \frac{\sigma^2}{2} = \sigma^2. \quad (2.1)$$

If we use the data to estimate this variance, with the usual sample variance formula, we get,

$$\begin{aligned} V &= \frac{1}{2} \left[\frac{\sum_{i=1}^2 (Y_{ai} - \bar{Y}_a)^2}{2-1} + \frac{\sum_{i=1}^2 (Y_{ri} - \bar{Y}_r)^2}{2-1} \right] \\ &= \frac{1}{2} \left[\frac{(Y_{a1} - Y_{a2})^2}{2} + \frac{(Y_{r1} - Y_{r2})^2}{2} \right]. \end{aligned} \quad (2.2)$$

Under our data generating mechanism, $Y_{zi} \stackrel{iid}{\sim} N(\mu_z, \sigma^2)$, the expected value of this variance estimate (2.2) matches the sampling variance (2.1). This is expected since the assumptions made in analysis match the data generating assumptions.

In order to illustrate the issues surrounding reference based analysis we now suppose one of the active arm patients deviates during the trial and their outcome, Y_{a2} , is missing. We consider first MCAR/MAR given treatment analysis then compare and contrast results with reference based analysis. The treatment effect of interest under the assumption of MCAR/MAR given treatment from a full ML analysis becomes, $\hat{\theta} = Y_{a1} - (Y_{r1} + Y_{r2})/2$. Under our assumed data generating mechanism this has expectation, $E[\hat{\theta}] = \mu_a - \mu_r$ and a long-run sampling variance,

$$\text{VAR}[\hat{\theta}] = \sigma^2 + \frac{\sigma^2}{2} = \frac{3\sigma^2}{2}. \quad (2.3)$$

We see in this simple example missing half the patients in the active arm directly inflates the sampling variance of the mean in the active arm by a factor of 2. The sampling variance of the treatment estimate increases, reflecting the loss of information, which is inevitable with missing data. Under the assumptions of our data generating mechanism the expected value of the variance estimator matches this long-run sampling variance. Thus also reflects the loss of information. The agreement between the sampling variance and variance estimator is again expected since the assumptions made in the analysis match the data generating assumptions.

When we do not see the data for the deviator in our sensitivity analysis we want to understand how different post-deviation behaviour effects the conclusions we draw. Suppose the deviating active patient switches treatment, so their outcome actually follows the same distribution as the patients in the reference arm, i.e. $Y_{a2} \stackrel{iid}{\sim} N(\mu_r, \sigma^2)$. We are still interested in the mean treatment group difference, where treatment is defined as randomised. This is our design based analysis.

If we were hypothetically able to see the deviator's outcome, under the assumptions of our updated data generating mechanism, the mean treatment group difference in expectation would now be $E[\hat{\theta}] = (\mu_a - \mu_r)/2$. The empirical long-run sampling variance for the treatment effect would be,

$$\text{VAR}[\hat{\theta}] = \frac{\sigma^2}{2} + \frac{\sigma^2}{2} = \sigma^2. \quad (2.4)$$

This is the sum of the variances of the averages of the two observations in each randomised treatment arm, which each individually have variance σ^2 . This is equivalent to the empirical long-run sampling variance for the treatment effect when patients are observed on-treatment (2.1). However, the expected value of the variance estimate, calculated for the original design based analysis using the sample variance formula under the assumptions of our updated data generating mechanism becomes,

$$E[V] = \frac{\sigma^2}{2} + \frac{\sigma^2}{2} + \frac{(\mu_a - \mu_r)^2}{4}. \quad (2.5)$$

The estimated variance is slightly higher than the sampling variance when we maintain our original design based analysis by randomised arm and the corresponding variance estimator. This is because the original design based analysis is no longer fully compatible with the updated data generating mechanism. (2.5) consequently incorporates an additional component that captures the heterogeneity within the active arm, a consequence of the deviator's off-treatment behaviour.

In practice when we do not see the deviator's outcome and use our observed reference group data to replace Y_{a2} by \tilde{Y}_{a2} where $E[\tilde{Y}_{a2}] = \mu_r$ the ML estimate of our treatment estimate is, $\hat{\theta} = (Y_{a1} + \tilde{Y}_{a2})/2 - (Y_{r1} + Y_{r2})/2$ which has expectation, $E[\hat{\theta}] = (\mu_a - \mu_r)/2$. This corresponds to assuming CR behaviour for our deviator, which in this simple scenario is equivalent to J2R and CIR.

The long-run sampling variance of this ML treatment estimate is then,

$$\text{VAR} [\hat{\theta}] = \frac{3}{4}\sigma^2 + \frac{1}{4}\text{VAR} [\tilde{Y}_{a2}] - \frac{1}{2}\text{COV} [\tilde{Y}_{a2}, Y_{r1} + Y_{r2}]. \quad (2.6)$$

By contrast, the variance estimator calculated from the design based analysis model using the sample variance formula is,

$$V = \frac{1}{2} \left[\frac{(Y_{a1} - \tilde{Y}_{a2})^2}{2} + \frac{(Y_{r1} - Y_{r2})^2}{2} \right].$$

This has expected value,

$$\begin{aligned} E[V] &= \frac{1}{2} \left[\frac{\sigma^2 + \text{VAR}(\tilde{Y}_{a2}) + (\mu_a - \mu_r)^2}{2} + \sigma^2 \right] \\ &= \frac{1}{2} \left[\frac{3\sigma^2}{2} + \frac{\text{VAR}(\tilde{Y}_{a2}) + (\mu_a - \mu_r)^2}{2} \right] \\ &= \frac{3\sigma^2}{4} + \frac{1}{4}\text{VAR}(\tilde{Y}_{a2}) + \frac{(\mu_a - \mu_r)^2}{4}. \end{aligned} \quad (2.7)$$

It is immediately clear that the long-run sampling variance of this ML treatment estimate (2.6) does not correspond with the variance estimator calculated from the primary design based analysis model (2.7). There are two key observations to note here. Firstly we see the variance estimator calculated from the primary design based analysis model (2.7) incorporates an additional component that captures the heterogeneity within the active arm, a consequence of the deviator's off-treatment behaviour, $(\mu_a - \mu_r)^2/4$. This is not surprising given the primary design based analysis model is inconsistent with the updated data generating mechanism. The sampling variance (2.6) does not incorporate this component. This is consistent with our earlier inferences when we supposed the CR deviation data was actually observable. This is a consequence of the partial mismatch between the assumptions of the analysis model and the updated data generation mechanism.

Moreover, when we use the observed data in the reference arm to create the imputation distribution for the active patient who copies reference, an unwanted covariance between these two naturally independent groups is introduced. We see this is included in the long-run sampling variance of our reference based treatment estimate (2.6) when we maintain our design based analysis by randomised arm. As we are examining the difference by randomised arm the unwanted covariance is taken away in the sampling variance. It is unwanted as when the active patient does actually CR behaviour in practice, they will be independent of all the other patients in the reference arm. This covariance term does not appear in the long-run sampling variance for the treatment estimate when the deviator is actually observed to CR (2.1).

As a consequence of borrowing information 'between' arms the long-run sampling variance of

our treatment estimate (2.6) becomes smaller than when the deviation data is actually observed under CR behaviour (2.4). Further the long-run sampling variance of our treatment estimate (2.6) becomes smaller than in our primary analysis (2.3) (assuming $(1/4)VAR[\tilde{Y}_{a2}] \leq (3/4)\sigma^2 + (1/2)COV[\tilde{Y}_{a2}, Y_{r1} + Y_{r2}]$), implying we have added information in the sensitivity analysis.

But this cannot be right for sensitivity analyses of the type we have proposed. We cannot have more information in the sensitivity analysis (2.6) than when the deviation data is actually observed under reference behaviour (2.5). A loss of information is inevitable with missing data. We must account for this. Also we must acknowledge the additional heterogeneity among the deviators.

This leads us to our first key principle that generally, *the expected value of the design based variance estimator if we were able to observe all the post-deviation data under the specified scenario (2.5) provides a natural lower bound for the expectation of the variance estimator in sensitivity analysis.* In sensitivity analysis our estimate of variance (2.5) must be scaled up to account for the missing information. We are not adding any additional information to the data in sensitivity analysis.

We now introduce a second trial setting with baseline and a single follow-up and apply our principle to this scenario to further highlight these issues in a more realistic setting. We develop this example in Chapter 3.

Application of principle

Consider a randomised trial with an active and reference arm ($z = a, r$), and two repeated measurements per patient i , (Y_{zi1}, Y_{zi2}) where time 1 is a baseline and time 2 a follow-up visit. Data from the patients in the reference arm are observed without deviation from the following bivariate normal distribution,

$$\begin{pmatrix} Y_{ri1} \\ Y_{ri2} \end{pmatrix} \sim N \left\{ \begin{pmatrix} \mu_1 \\ \mu_{r2} \end{pmatrix}; \begin{bmatrix} \sigma_{r11} & \sigma_{r12} \\ \sigma_{r12} & \sigma_{r22} \end{bmatrix} \right\}, i = 1, \dots, n_r.$$

In the active arm post-baseline n_d of the n_a patients deviate, n_o of the n_a active patients do not such that $n_o + n_d = n_a$. Let \mathcal{D} and \mathcal{O} define the sets of indices for the patients in the active arm who do and do not deviate respectively. Data from the active arm comes from the following bivariate normal distributions,

$$\begin{pmatrix} Y_{ai1} \\ Y_{ai2} \end{pmatrix} \sim N \left\{ \begin{pmatrix} \mu_1 \\ \mu_{a2} \end{pmatrix}; \begin{bmatrix} \sigma_{a11} & \sigma_{a12} \\ \sigma_{a12} & \sigma_{a22} \end{bmatrix} \right\}, i \in \mathcal{O}$$

$$\begin{pmatrix} Y_{ai1} \\ Y_{ai2} \end{pmatrix} \sim N \left\{ \begin{pmatrix} \mu_1 \\ \mu_{d2} \end{pmatrix}; \begin{bmatrix} \sigma_{d11} & \sigma_{d12} \\ \sigma_{d12} & \sigma_{d22} \end{bmatrix} \right\}, i \in \mathcal{D}.$$

The design based analysis is the unadjusted difference in means between the two randomised arms

at time 2. In Chapter 3 we develop this example for the baseline adjusted treatment effect. We first suppose that post-deviation data is observed. Interim steps in the below calculations are presented in Appendix B.1. The unadjusted treatment effect of interest is computed as,

$$\hat{\theta} = \frac{1}{n_a} \left(\sum_{i \in \mathcal{O}} Y_{ai2} + \sum_{i \in \mathcal{D}} Y_{ai2} \right) - \frac{1}{n_r} \sum_{i=1}^{n_r} Y_{ri2}. \quad (2.8)$$

Conditioning on n_a , n_d and n_r , the expected treatment effect at visit 2 is,

$$E[\hat{\theta}] = \left(\frac{n_o}{n_a} \mu_{a2} + \frac{n_d}{n_a} \mu_{d2} \right) - \mu_{r2}.$$

The variance of the treatment effect is estimated by,

$$V = \frac{\hat{\sigma}_{r22}}{n_r} + \frac{\hat{\sigma}_{a22}}{n_a} = \frac{1}{n_r-1} \sum_{i=1}^{n_r} (Y_{ri2} - \bar{Y}_{r2})^2 + \frac{1}{n_a-1} \sum_{i=1}^{n_a} \left(Y_{ai2} - \frac{n_o}{n_a} \bar{Y}_{a2,o} - \frac{n_d}{n_a} \bar{Y}_{a2,d} \right)^2$$

which has expected value,

$$E[V] = \frac{\sigma_{r22}}{n_r} + \frac{\frac{n_o}{n_a} \sigma_{a22} + \frac{n_d}{n_a} \sigma_{d22}}{n_a} + \frac{n_o n_d \Delta^2}{(n_a - 1) n_a^2} \quad (2.9)$$

where $\Delta = \mu_{a2} - \mu_{d2}$. This is asymptotically equivalent to,

$$E[V] = \frac{\sigma_{r22}}{n_r} + \frac{\frac{n_o}{n_a} \sigma_{a22} + \frac{n_d}{n_a} \sigma_{d22}}{n_a} + \frac{n_o n_d \Delta^2}{n_a^3}.$$

Assuming the variance is equal in both arms, $\sigma_{r22} = \sigma_{d22} = \sigma_{a22} = \sigma_{22}$ and $n_r = n_a = n$ this is equivalent to,

$$E[V] = \frac{2\sigma_{22}}{n} + \frac{\Delta^2 n_d n_o}{n^3}. \quad (2.10)$$

As discussed the expected value of the design based variance estimator if we were able to observe all the post-deviation data under the specified scenario provides a natural lower bound for the expectation of the variance estimator in sensitivity analysis.

In the presence of missing post-deviation data we require an estimate of variance for the imputation treatment effect that is larger than the specified lower bound (2.10) in expectation. This is to take into account the uncertainty created by the loss of information. Loss of efficiency is an inevitable consequence of missing data.

Conventional long-run sampling variance

Our earlier simple example indicated that the conventional long-run sampling variance of the treatment effect will be too small in reference based settings, since a negative covariance between the deviators and reference group is induced. We now explore this further. Suppose post-deviation data are now missing and our treatment effect (2.8) is estimated by ML assuming CR behaviour for the deviators. In this setting, where time 1 is a baseline, this is equivalent to assuming J2R or CIR. The estimated treatment effect and the empirical long-run sampling variance of the treatment effect obtained using CR MI will be asymptotically equivalent to that of the ML estimator investigated here. Denoting the ML estimates of μ_{a2} as $\hat{\mu}_{a2}$ and μ_{r2} and μ_{d2} as $\hat{\mu}_{r2}$ the ML estimate of the treatment effect will be,

$$\hat{\theta}_{ML} = \frac{n_o}{n_a} \hat{\mu}_{a2} + \frac{n_d}{n_a} \hat{\mu}_{r2} - \hat{\mu}_{r2} = \frac{n_o}{n_a} (\hat{\mu}_{a2} - \hat{\mu}_{r2}).$$

We see that as $n_d \rightarrow n_a$ and $n_o \rightarrow 0$ the sampling variance of this estimate will tend to zero. The sampling variance will indeed be less than the variance computed using the data (2.10), unless $n_d = 0$ where $n_o = n_a$.

Thus the conventional long-run sampling variance is not what we want for exploring the sensitivity of the primary analysis estimator as assumptions vary treatment group behaviour. It is too small partly because we have used the observed reference data twice. We have said the active deviators mean is 100% correlated with the mean in the reference arm. There is also the partial mismatch between the data generation mechanism and analysis model. The design based analysis model does not make the same assumptions as the imputation model.

The conventional long-run sampling variance is therefore misleading. It does not incorporate any uncertainty. With missing post-deviation data we require an estimate of variance that is larger than (2.10), to properly reflect the uncertainty associated with the unknown observations.

2.2.3 Rubin's treatment and variance estimate

Rubin's treatment estimator is not anticipated to be biased since each imputation only adds random terms with an expectation of zero to the observed data quantities. The average of the treatment estimates over the imputed data sets is then computed. The MI and ML reference based estimators will be equivalent.

We now consider Rubin's variance estimate in the reference based MI setting. As discussed by Meng in 1994 [46] Rubin's variance estimator provides valid inference for the long-run sampling variance

of the MI estimator when imputations are repeated draws from a Bayesian prediction model, and the procedure for analysing the multiply imputed datasets is congenial to the imputation model.

In the MAR setting, the primary design based analysis model that categorises patients as randomised is congenial or consistent with the MAR data generating mechanism that imputes data for patients as randomised. In this case Rubin’s variance formula provides correct frequentist inference when we are looking for the long-run sampling variance of the MI estimator, which asymptotically matches the long-run variance of the ML estimator. This would initially imply Rubin’s variance estimator may not be what is required in the reference based settings, since as discussed we do not want the long-run sampling variance of the ML estimator; or that of the MI estimator of the reference based estimand (as the MI and ML reference based estimator will be equivalent). This violates our principle.

However, consider the J2R setting where post-deviation, patients’ mean response distributions in the active group follow that observed for a reference group. There is no single Bayesian (imputation) model whose posterior mean of the treatment effect θ and corresponding posterior variance will be asymptotically the same as the estimate and variance from the primary analysis procedure which categorises patients as randomised. In this case and in the other reference based cases there is a partial mismatch between the data generation model and the analysis model. The imputation and analysis model are not congenial, violating the frequency validity of Rubin’s rules long-run.

In this particular context where the imputer is assuming more in the imputation model Meng [46] showed that due to the additional information built into the imputation model, Rubin’s variance estimator will actually over estimate the long-run sampling variance. We therefore expect that Rubin’s variance estimator will be larger than long-run sampling variance in the reference based settings. The key question of interest is, does it give us what we need? Does it fulfil our general principle?

We summarise the key points so far regarding variance estimation in the reference based cases:

1. The expected value of the design based variance estimator if we observed post-deviation data provides a lower bound for the expectation of the variance estimator in sensitivity analysis.
2. An estimate of variance that is above this bound is required to take into account the loss of information.
3. The conventional long-run variance of the ML estimator is not appropriate in the reference based settings, it is too small.
4. It is unclear whether Rubin’s MI variance formula provides us with a suitable estimate of variance.

In the following section we investigate the statistical performance of Rubin’s MI treatment estimator and variance estimator alongside the conventional long-run sampling variance of the reference based MI/ML treatment estimator in a more complex longitudinal setting through simulation.

2.3 Exploratory simulation study

A simulation study was conducted to evaluate the performance of Rubin's MI treatment and variance estimator for each of the proposed imputation options of Carpenter, Roger and Kenward [13]; MAR, CR, J2R, CIR and LMCF.

2.3.1 Methods

The data from the asthma trial (see Section 1.7.2) formed the motivating example for generating the datasets [54]. The primary outcome was forced expiratory volume, FEV₁ (in Litres), recorded at baseline, 2, 4, 8 and 12 weeks, however completion rates at week 12 were only 41% and 79% in the placebo and active arms respectively. A summary of the observed deviation patterns can be found in Section 1.7.2. Data (baseline, 2, 4, 8 and 12 week FEV₁) were generated using an underlying MVN distribution with means and covariance matrices obtained in the study as,

$$\mu_{\text{placebo}} = [2.06, 1.97, 1.94, 1.91, 1.88], \mu_{\text{active}} = [2.05, 2.17, 2.21, 2.22, 2.20]$$

$$\Sigma_{\text{placebo}} = \begin{bmatrix} 0.35 & 0.29 & 0.22 & 0.26 & 0.17 \\ 0.29 & 0.45 & 0.28 & 0.35 & 0.41 \\ 0.22 & 0.28 & 0.33 & 0.26 & 0.24 \\ 0.26 & 0.35 & 0.26 & 0.45 & 0.37 \\ 0.17 & 0.41 & 0.24 & 0.37 & 0.54 \end{bmatrix}$$

$$\Sigma_{\text{active}} = \begin{bmatrix} 0.42 & 0.42 & 0.43 & 0.43 & 0.39 \\ 0.42 & 0.56 & 0.53 & 0.55 & 0.51 \\ 0.43 & 0.53 & 0.64 & 0.59 & 0.54 \\ 0.43 & 0.55 & 0.59 & 0.70 & 0.60 \\ 0.39 & 0.51 & 0.54 & 0.60 & 0.60 \end{bmatrix}$$

One thousand independent complete datasets were generated in total and a sample size of n=180 patients (n=90 placebo, n=90 active) was assumed for all simulations.

A monotone missingness mechanism, dependent on treatment group, baseline FEV₁ and the previously observed outcome was imposed on the FEV₁ outcomes at 4, 8 and 12 weeks using logistic regression models on indicators of missingness. This ensured the missingness was missing at random (MAR). Let $i = 1, \dots, n$ index patients and $j = 0, 1, 2, 3, 4$ index the planned observation times of baseline, 2, 4, 8 and 12 weeks respectively. Let the outcome for each patient i , at time j be denoted by y_{ij} and the response indicator be defined as $R_{ij} = 1$ if y_{ij} is observed without deviation, otherwise $R_{ij} = 0$. The model for response, where z_i denotes the treatment group assignment for patient i is denoted,

$$\begin{cases} \text{logit } Pr(R_{ij} = 1 | y_{i1}, \dots, y_{ij}, z_i) = \alpha_0 + \alpha_1 z_i + \alpha_2 y_{i0} + \alpha_3 y_{i(j-1)}, j = 2, 3, 4 \\ Pr(R_{ij} = 1) = 0 \text{ if } R_{i(j-1)} = 0. \end{cases}$$

Initially all α 's were obtained from the information in the observed asthma study data to give an overall rate of missingness of approximately 40% at week 12 (59% placebo arm, 21% active arm) with dropout pattern as observed in the trial. Three other settings were considered, with overall rates of missingness of approximately 30%, 20% and 10% at week 12. This was achieved by calibrating the value of α_0 in the missingness model accordingly. The exact values of the used α 's are presented in Table A.1. In each case a probability of response for each incomplete outcome was calculated for every patient and compared against a random value from the Uniform[0, 1] distribution. The outcome was set to missing if their uniformly distributed value exceeded the calculated probability, or if the outcome at the previous follow-up time point was unobserved.

The reference based MI approach of Carpenter, Roger and Kenward [13] was applied to each incomplete dataset under MAR, CR, J2R, CIR and LMCF with 1000 imputations. For the methods alluding to a reference group, the analysis was conducted once with placebo as reference and second with active as reference. We used the same set of 1000 incomplete datasets for each imputation procedure and setting to eliminate sample variability and strengthen comparisons.

In each scenario and simulation run post-deviation data were regenerated using the appropriate conditional normal distributions for the underlying assumption, to compute the treatment effect and its SE had the post-deviation data been observed (averaged over the 1000 simulations for comparisons).

The analysis model was a linear regression of the 12 week response on treatment group, adjusted for baseline. The main outcomes of interest in each scenario were Rubin's MI estimate and Standard Error (SE) of the treatment effect (averaged over the 1000 simulations for comparison), and the empirical long-run SE of the MI estimator over the 1000 simulations.

2.3.2 Results

Table 2.2 summarises the results for each of the imputation procedures in the 40% missingness setting, which mimics the deviation rate observed in the asthma study. True θ is the true value of the treatment effect; $\hat{\theta}$ (MCSE range 0.002 to 0.003) is the average of the completed-data treatment effect estimates, thus represents the average treatment effect obtained when the deviators are actually observed to have the indicated post-deviation behaviour; Rubin's $\hat{\theta}_{MI}$ (MCSE range 0.002 to 0.003) is the average treatment effect using MI under the indicated assumption with Rubin's Rules. We expect these to be equal.

In all scenarios, as expected, there is minimal to no bias of Rubin's treatment effect, Rubin's $\hat{\theta}_{MI}$. Both $\hat{\theta}$ and Rubin's $\hat{\theta}_{MI}$ lie within 2 monte-carlo-standard errors (MCSE) of the True θ in all cases.

Model SE (MCSE range 0.0001 to 0.0002) is the average of the completed-data model based standard errors i.e. the average of the model based standard errors we would obtained if the

deviators are actually observed; Empirical SE is the empirical long-run standard error of the 1000 completed-data treatment estimates (MCSE range 0.0018 to 0.0021); Rubin’s SE_{MI} (MCSE range 0.00018 to 0.00022) is the square root of the average estimate of variance obtained using MI with Rubin’s rules; Empirical SE_{MI} (MCSE range 0.001 to 0.002) is the empirical long-run standard error of the 1000 MI treatment effects, which is asymptotically equivalent to that of the ML estimator since $\hat{\theta}_{MI}$ is unbiased, and COV Rubin’s MI is the coverage of 95% confidence interval from Rubin’s rules (MCSE range 0.42 to 0.69).

The average of the completed-data model based standard errors, Model SE, which approaches the expected value of the design based variance where all post-deviation data is observed under the specified scenario, provides a lower bound for the variance of the treatment effect. With missing post-deviation data we expect a larger variance estimate to reflect the introduced uncertainty. We see in all cases Rubin’s estimate of the standard error, Rubin’s SE_{MI} , is larger than this lower bound, as required.

Under MAR we see Rubin’s estimate of the standard error of the treatment effect estimates the long-run sampling standard error, Empirical SE_{MI} , without bias (within $2 \times MCSE$). Nominal coverage is achieved (95%) as expected.

For all reference based methods, except CR-placebo, the empirical standard error of the MI imputation estimator, Empirical SE_{MI} , whose value is asymptotically equal to the long-run variance of the ML estimator is lower than that based on the complete data. This is as expected since the long-run MI/ML variance estimator includes the induced negative covariance between the deviators and reference group. The variance in treatment contrasts decreases as the sample becomes more homogeneous, whereas the design based estimated variance does not.

Thus for all reference based imputation options it is evident that Rubin’s MI estimate of the standard error is desirably overestimating the long-run variability of the MI/ML point estimate, as expected. The coverage of 95% confidence intervals from Rubin’s rules is consequently conservative, ranging from 0.955 to 0.999. This overestimation is less noticeable for CR-placebo due to the lower number of active deviators. For the procedures that assume statistical behaviour of a reference arm post-deviation, the difference between Rubin’s MI variance estimator and the long-run empirical variance is larger where the active arm is the reference as overall there is a greater proportion of patients deviating with reference to the other arm.

Figures 2.6 and 2.7 summarise the results for MAR and the reference based settings with 10–40% missingness. In all missingness scenarios we see Rubin’s estimate of variance is greater than the lower bound (variance estimate if the deviators could be observed) and increases as the proportion of missing data increases. This is in contrast to the long-run empirical sampling variance for the MI/ML estimator which is lower than the variance estimator from the complete data and decreases towards zero as the amount of missing data increases. The long-run sampling variance, as anticipated, is misleading for sensitivity analysis as it does not incorporate the additional variability incurred with the greater loss of data.

Method	True θ		Post-deviation data complete			Post-deviation data missing		
	$\hat{\theta}$	Model SE	Empirical SE	Rubin's $\hat{\theta}_{MI}$	Rubin's SE_{MI}	Empirical SE_{MI}	COV	Rubin's MI
MAR	0.336	0.332	0.089	0.091	0.334	0.102	0.102	0.950
J2R Placebo	0.261	0.263	0.093	0.095	0.265	0.104	0.085	0.982
J2R Active	0.145	0.142	0.081	0.083	0.143	0.108	0.066	0.999
CR Placebo	0.334	0.331	0.091	0.093	0.333	0.101	0.097	0.955
CR Active	0.277	0.272	0.080	0.082	0.273	0.091	0.074	0.981
CIR Placebo	0.324	0.322	0.091	0.093	0.324	0.102	0.092	0.964
CIR Active	0.298	0.294	0.080	0.082	0.296	0.091	0.075	0.976
LMCF	0.299	0.295	0.089	0.091	0.298	0.099	0.078	0.980

Table 2.2: Performance of Rubin's Rules in a simulation study with 40% missingness; θ = treatment effect, $\hat{\theta}$ = average of the completed-data treatment estimates, Model SE = average of the completed-data model based standard errors, Empirical SE = standard error of the completed-data treatment estimates over 1000 simulations, Rubin's θ_{MI} = average of Rubin's MI treatment estimates, Rubin's SE = square root of the average estimate of variance obtained using Rubin's rules, Empirical SE_{MI} = standard error of Rubin's MI treatment estimates over 1000 simulations, COV Rubin's MI = coverage of 95% confidence interval from Rubin's rules.

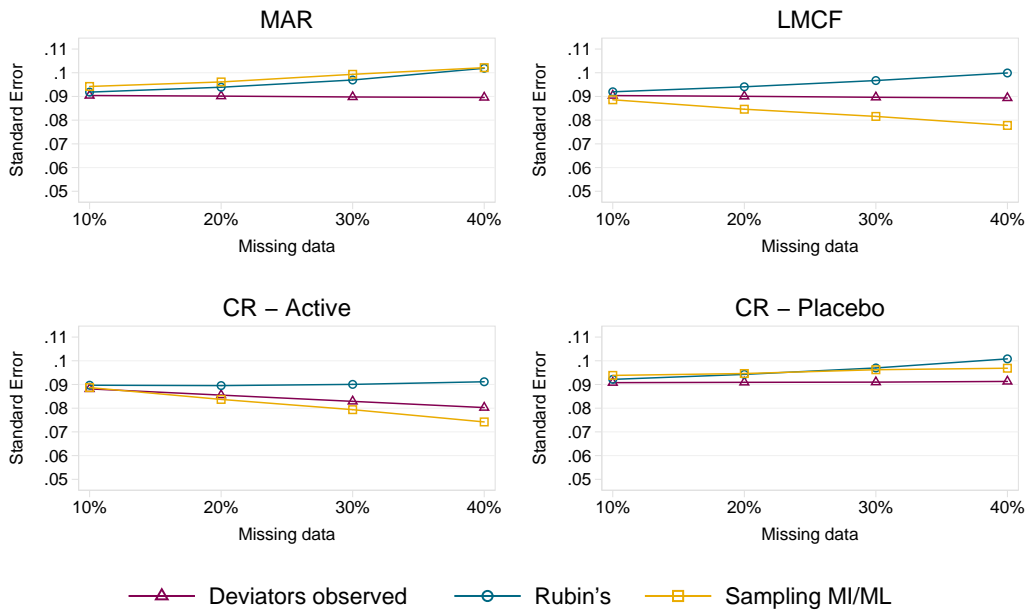


Figure 2.6: Rubin's SE (Mean Rubin's SE_{MI}) vs. conventional long-run sampling variance of the MI/ML estimate (SE of the 1000 MI estimates, Empirical SE_{MI}) vs. SE where deviators observed (Mean Model SE over 1000 simulations) for MAR, LMCF and CR.

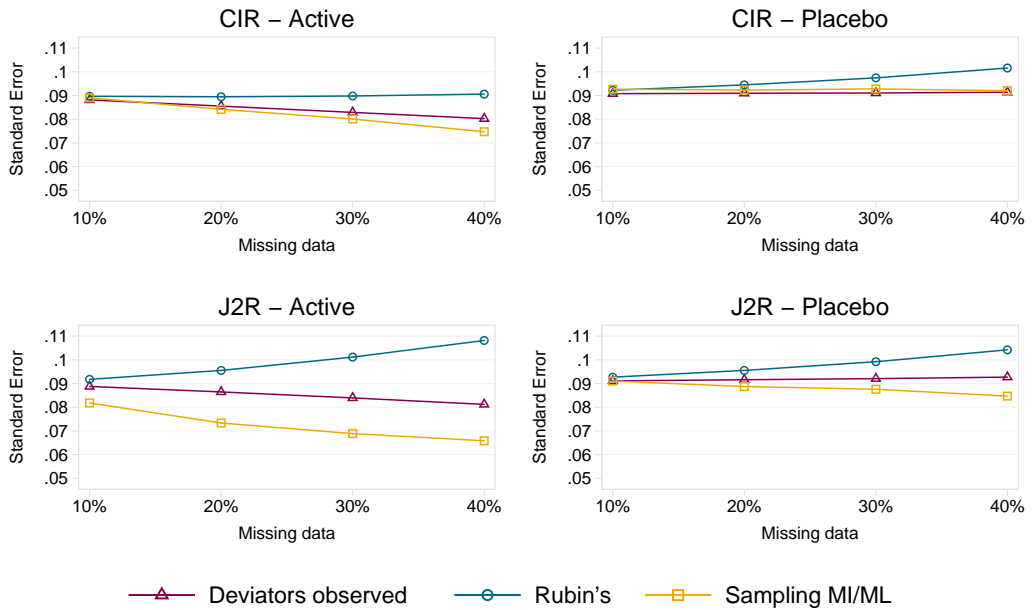


Figure 2.7: Rubin's SE (Mean Rubin's SE_{MI}) vs. conventional long-run sampling variance of the MI/ML estimate (SE of the 1000 MI estimates, Empirical SE_{MI}) vs. SE where deviators observed (Mean Model SE over 1000 simulations) for CIR and J2R.

2.3.3 Discussion

Rubin's MI treatment estimator is unbiased for the reference based treatment estimate. The long-run sampling variance of the reference based MI treatment effect, which is equivalent to the long-run sampling variance of the ML treatment effect, is confirmed as being too small. Rubin's

MI variance estimate as expected is different to this and has desirable properties. Rubin's MI variance estimate is larger than the design based variance where the deviation data is observed, which provides a lower bound for the variance of the treatment effect. Rubin's variance estimate also increases as the proportion of missing data increases.

2.4 Summary

Rubin's MI treatment estimate is confirmed as unbiased for the reference based treatment estimator. We therefore do not explore this issue further in the subsequent developments.

With missing data in reference based sensitivity analysis we require an estimate of variance that is greater than the estimated variance we would observe from the design based analysis under the given scenario if all data were actually observed. The conventional long-run sampling variance of the reference based treatment effect (estimated via MI or ML) does not meet this criteria. Rubin's MI variance estimator does satisfy this principle and has desirable properties. It is larger than the design based variance estimate we would obtain were we able to observe the reference based post-deviation data, and increases as the proportion of missing data increases. Thus MI with Rubin's rules appears to be a fruitful route for performing reference based analyses.

However further exploration is required in each reference based setting to assess exactly what Rubin's variance estimator is giving us. In the next chapter we discuss and define desirable properties of the required variance estimator in the reference based setting and compare this to Rubin's estimator with baseline and a single follow-up. In this setting the exact properties of Rubin's variance estimate can be derived. This paves the way for more generalised results on the properties of Rubin's variance estimator in reference based MI in Chapter 4.

Chapter 3

Behaviour of Rubin's variance estimator in reference based multiple imputation; baseline and single follow-up

In this chapter we define and derive the required variance estimate for the reference based treatment estimator in Section 3.1. We then explore whether Rubin's MI variance estimator provides an acceptable estimate of this in the baseline and single follow-up trial setting. First, for the unadjusted treatment estimator in Section 3.2. Following theoretical results, a supporting simulation study is presented in Section 3.3. We then extend our inferences for the baseline adjusted treatment estimator in Section 3.4. We summarise our findings in Section 3.5. This leads to a more general proposition and proof on the properties of Rubin's variance estimator in the following chapter.

Throughout this chapter we retain the notation introduced in Chapter 2 where $\hat{\theta}$ denotes the estimated treatment effect of interest from the design based analysis. The estimated variance of the treatment effect is denoted by V . As introduced in Subsection 1.3.5, a de-jure estimand compares the effect of treatments as randomised that are taken strictly according to the protocol. It does not necessarily require the MAR assumption in the presence of missing data. However as these two are linked throughout this chapter our de-jure estimate, indexed with the subscript $_{DJ}$ denotes analysis under MAR. A de-facto estimand compares the effect of treatments as randomised, regardless of protocol adherence. A de-facto estimate, indexed with the subscript $_{DF}$, therefore denotes a general reference based analysis, that is under one of CR, J2R, CIR or LMCF. Where we refer to a specific de-facto analysis only we index the estimate with the subscript $_{CR, J2R, CIR \text{ or } LMCF}$ as appropriate.

Setting

We consider the same two arm trial setting introduced in Subsection 2.2.2, with the additional assumption that the variance-covariance matrix is the same in both treatment arms. This is a trial

of an active versus reference treatment, denoted by $z = a, r$, with $i = 1, \dots, n$ subjects per arm with a baseline (time 1) and single follow-up outcome (time 2). The unadjusted difference in means between the two trial arms at time 2 is the key parameter of interest. In Section 3.4 we develop this example and consider the baseline adjusted difference in means as the key parameter of interest. All of the reference arm ($z = r$) and n_o of the n active arm patients ($z = a$) are observed at both times without deviation. Post-baseline the remaining n_d active arm patients deviate such that $n_o + n_d = n$. We condition on n_d and where \mathcal{O} and \mathcal{D} define the sets of indices for the completers and deviators in the active arm respectively, assume the data comes from the following bivariate normal distributions with equal variance-covariance structures,

$$\begin{aligned} \begin{pmatrix} Y_{zi1} \\ Y_{zi2} \end{pmatrix} &\sim N \left\{ \begin{pmatrix} \mu_{z1} \\ \mu_{z2} \end{pmatrix}; \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \right\}, \quad i = 1, \dots, n \text{ for } z \in r \text{ or } i \in \mathcal{O} \text{ for } z \in a \\ \\ \begin{pmatrix} Y_{zi1} \\ Y_{zi2} \end{pmatrix} &\sim N \left\{ \begin{pmatrix} \mu_{z1} \\ \mu_{d2} \end{pmatrix}; \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \right\}, \quad i \in \mathcal{D} \text{ and } z \in a. \end{aligned}$$

As described in Subsection 2.2.2, the expected value of the variance estimator for the treatment effect from the design based analysis, if we observe all the post-deviation data under the specified de-facto scenario (one of CR, J2R, CIR or LMCF), provides a lower bound for the expectation of the appropriate variance estimator in the presence of missing data. We now denote this lower bound by $E[V_{\text{DF, full}}]$ and in the current setting with equal known variance, as derived in Subsection 2.2.2 (2.10) this is,

$$E[V_{\text{DF, full}}] = \frac{2\sigma_{22}}{n} + \frac{\Delta^2 n_d n_o}{n^3}, \quad (3.1)$$

where $\Delta = \mu_{a2} - \mu_{d2}$ and μ_{d2} represents the proposed mean at time 2 under the particular de-facto scenario of interest. In the presence of missing data we require an estimate of variance whose expectation is greater than this and appropriately incorporates the loss of information. So the first question is, *what is the appropriate inflation factor to apply to (3.1) with missing data?* The second question is, *does Rubin's MI variance estimator give an acceptable estimate of the desired variance estimator constructed from (3.1) and the appropriate inflation factor?*

3.1 Information anchored sensitivity analysis

We focus here on the appropriate inflation factor to apply to (3.1) in the presence of missing data when we use reference based MI for sensitivity analysis in the way set out in Chapter 2.

A natural principle for sensitivity analyses which retain the primary design based analysis model is to keep the information loss due to missing data constant or *anchored* across the primary and sensitivity analyses. A loss of information is inescapable with missing data and the extent of information loss in the primary analysis under MAR is linked to the chosen analysis model, rather than the actual proportion of incomplete records. Thus we propose the loss of information due to

missing data in the primary analysis should equal the loss of information due to missing data in all sensitivity analyses which retain the design based analysis model.

The variance of our treatment estimator in reference based MI will be derived from the variance of the imputed data. So ultimately is in the control of the analyst through the assumption about the variability of the missing data. Under reference based MI we could alternatively inflate or deflate the variance of the missing data. A wealth of potential options exist, but these all require external judgements beyond the observed trial data. Essentially arbitrary choices. Anchoring the information loss in the primary analysis provides context and a natural logical foundation.

When conducting reference based sensitivity analysis we should not be artificially adding any extra information to our data. Sensitivity analysis should not inject information ‘by the back door.’ Neither do we want to loose valuable information collected in the trial. In summary we want to preserve the loss of information in the primary design based analysis in the design based sensitivity analysis.

In reference based sensitivity analysis the primary analysis is conducted under the MAR assumption. As discussed in Section 1.3 analysis under MAR is a natural starting point since it is often most plausible that patients who deviate will exhibit the same response as those with the same history still in the trial i.e. those in the same treatment arm. Further modelling of the dropout in the reference based settings is not required. The primary MAR analysis provides the building blocks for the sensitivity analysis. Our post-deviation data distributions are pieced together using the parameters from the underlying MAR model. We can therefore measure the information loss in the primary MAR analysis, relative to when there is no missing data. Then explore how this relates to the inflation in Rubin’s variance estimator under reference based post-deviation behaviour, when the deviation data is missing, relative to when there is no missing data.

Let $V_{\text{DJ, full}}$ denote the variance estimator for the treatment effect from the design based analysis with full data i.e. no deviations. When deviations occur and post-deviation data is unobserved let $V_{\text{DJ, MI}}$ denote Rubin’s variance estimator for the treatment effect from the design based MAR MI analysis (primary analysis). We propose that the variance estimator required for the treatment estimate in sensitivity analysis denoted by V_{anchored} is,

$$V_{\text{anchored}} = \frac{V_{\text{DJ, MI}}}{V_{\text{DJ, full}}} \times V_{\text{DF, full}}. \quad (3.2)$$

$V_{\text{DJ, MI}}/V_{\text{DJ, full}}$ provides a measure of the information loss in the primary analysis. When this inflation factor is applied to $V_{\text{DF, full}}$ —the design based variance estimator if we observed all the reference based post-deviation data— this gives our proposed information anchored variance estimate.

We recall here that the information on a parameter, denoted by I , is defined as the reciprocal of its variance. If the variance is small, the information on the associated parameter is thus large, however if the variance is large, the information on the parameter will be small. We can therefore also express the proposed variance as,

$$V_{\text{anchored}} = \frac{I_{\text{DJ, full}}}{I_{\text{DJ, MI}}} \times V_{\text{DF, full}}.$$

The information preserving principle described by Rubin [19] tells us, $I_{\text{obs}} = I_{\text{full}} - I_{\text{miss}}$, that is $I_{\text{obs}} + I_{\text{miss}} = I_{\text{full}}$, i.e. the observed information plus the missing information equals the full information. We can therefore immediately see that this inflation directly takes account and preserves the fraction of missing information in the primary design based analysis. The inflation is a measure of the loss of precision due to non-response as follows,

$$\frac{I_{\text{DJ, full}}}{I_{\text{DJ, MI}}} = \frac{I_{\text{DJ, MI}} + I_{\text{miss}}}{I_{\text{DJ, MI}}} = 1 + \frac{I_{\text{miss}}}{I_{\text{DJ, MI}}}.$$

We define the proposed estimate of variance that keeps the information loss constant across the primary and sensitivity analysis as the *information anchored variance*.

Example

In the current setting outlined at the beginning of this chapter, the expected value of the design based variance estimator in full data, if no deviations occur (i.e. everyone on-treatment) using standard normal distribution results is,

$$E[V_{\text{DJ, full}}] = \frac{2\sigma_{22}}{n}.$$

To establish the required information anchored variance in sensitivity analysis we need to compute the expected value of the variance estimator for the treatment effect in the primary design based MAR setting. When performing MI under MAR and the congeniality condition, Rubin's variance estimator, $V_{\text{DJ, MI}}$, yields appropriate inference that is equivalent (as $K \rightarrow \infty$) to the corresponding full likelihood analysis. We now analytically derive the expectation of Rubin's MI variance estimator under MAR which demonstrates this. The result of the derivation is then used in the subsequent exploration of the reference based MI variance estimator.

3.1.1 Rubin's variance estimator under MAR

To derive the expectation of Rubin's MI variance estimator for the treatment estimate under MAR an appropriate imputation distribution for the missing outcomes at time 2 is required, with a suitable posterior for the included parameters. We assume the data is normal. In practice, as detailed in Section 2.1 an improper prior for the parameters of the imputation distribution and an uninformative prior for the covariance matrix is assumed. Throughout this exposition and the subsequent reference based expositions the variance is assumed to be known i.e. we do not incorporate a prior distribution on the variance in imputation without any serious loss of generality. This is following inference by Carpenter and Kenward in [44] (p. 56–59) that examines Rubin's rules for a simple mean with known and unknown variance. Carpenter and Kenward show, because of the conditional argument underlying the derivation, the additional complication of an unknown variance does not have a profound effect on the development. It simply results in multiples of terms that are ≈ 1 and thus approach 1 fast as $n \rightarrow \infty$. The results are asymptotically the same.

To simplify subsequent calculations we define $Y_{ai1}^* = Y_{ai1} - \bar{Y}_{a1,o}$ where $\bar{Y}_{a1,o} = \frac{1}{n_o} \sum_{i \in \mathcal{O}} Y_{ai1}$. The data from the completers in the active arm at time 1 is denoted by the column vector,

$$\mathbf{Y}_{\mathbf{a}1,o}^* = \{Y_{ai1}^*; i \in \mathcal{O}\}^T.$$

The observed data at time 2 in the active arm is denoted by the column vector $\mathbf{Y}_{\mathbf{a}2,o}$ where,

$$\mathbf{Y}_{\mathbf{a}2,o} = \{Y_{ai2}; i \in \mathcal{O}\}^T.$$

Under MAR, our imputation model is formed from the regression of Y_{ai2} on Y_{ai1}^* for the set of completers in the active arm ($i \in \mathcal{O}$). Let $\mathbf{x}_i = (1, Y_{ai1}^*)^T$ then the conditional model of interest for imputation is,

$$Y_{ai2} = \mathbf{x}_i^T \beta + e_i, \quad e_i \sim N(0, \sigma_{2.1}) \quad i = 1, \dots, n_o,$$

where $\sigma_{2.1}$ is the residual variance from the regression of Y_{ai2} on Y_{ai1}^* . Re-expressed in matrix notation the conditional distribution of interest $\mathbf{Y}_{\mathbf{a}2,o} | \mathbf{Y}_{\mathbf{a}1,o}^*$ is,

$$(\mathbf{Y}_{\mathbf{a}2,o} | \mathbf{Y}_{\mathbf{a}1,o}^*) \sim N(\mathbf{X}\beta; \sigma_{2.1}),$$

where \mathbf{X} is the design matrix,

$$\mathbf{X} = \begin{bmatrix} 1 & Y_{a11}^* \\ 1 & \cdot \\ 1 & Y_{an_o1}^* \end{bmatrix}.$$

The ordinary least squares estimates of the parameters, β , of this distribution which we denote by $\hat{\beta}$, are obtained by fitting the conditional model to the observed data and calculating $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}_{\mathbf{a}2, \mathbf{o}}$ which gives,

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_{20.1} \\ \hat{\beta}_{21.1} \end{pmatrix} = \begin{pmatrix} \bar{Y}_{a2,o} \\ r/q \end{pmatrix},$$

where $\bar{Y}_{a2,o} = \frac{1}{n_o} \sum_{i \in \mathcal{O}} Y_{ai2}$, $r = \sum_{i \in \mathcal{O}} (Y_{ai1} - \bar{Y}_{a1,o}) (Y_{ai2} - \bar{Y}_{a2,o})$ and $q = \sum_{i \in \mathcal{O}} (Y_{ai1} - \bar{Y}_{a1,o})^2$, with associated variance $\sigma_{2.1} (\mathbf{X}^T \mathbf{X})^{-1}$, which is,

$$\text{Var}(\hat{\beta}) = \sigma_{2.1} \begin{pmatrix} n_o^{-1} & 0 \\ 0 & q^{-1} \end{pmatrix},$$

for $\sigma_{2.1} = \sigma_{22} - \sigma_{12}^2 / \sigma_{11}$. We assume the data is normal and improper priors for the missing observations, so the posterior of the model parameters (β) is approximately normal and let the observed data dominate the prior. The variance is assumed to be known without any serious loss of generality, following inference by Carpenter and Kenward [44]. Thus the MAR imputation model for patient i and imputation k is,

$$\tilde{Y}_{ai2,k} = \bar{Y}_{a2o} + u_k + (r/q + b_k) (Y_{ai1} - \bar{Y}_{a1o}) + e_{i,k}, \text{ for } i \in \mathcal{D}, k = 1, \dots, K, \quad (3.3)$$

where,

$$\begin{aligned} u_k &\sim N(0, n_o^{-1} \sigma_{2.1}) \\ b_k &\sim N(0, q^{-1} \sigma_{2.1}) \\ e_{i,k} &\sim N(0, \sigma_{2.1}). \end{aligned}$$

Following MI from the above model for the deviators, we are interested in establishing the expectation of Rubin's MI estimator and variance estimator for the treatment difference at time 2, which we denote respectively as $\hat{\theta}_{\text{DJ, MI}}$ and $V_{\text{DJ, MI}}$. Firstly,

$$E(\hat{\theta}_{\text{DJ, MI}}) = E\left(\frac{1}{K} \sum_{k=1}^K \hat{\theta}_{\text{DJ, } k}\right).$$

Where the sample mean treatment difference, at time 2 in the k^{th} imputation set, $\hat{\theta}_{\text{DJ, } k}$ is obtained as,

$$\hat{\theta}_{\text{DJ, } k} = \hat{\mu}_{a2,k} - \hat{\mu}_{r2} = \bar{Y}_{a2,k} - \bar{Y}_{r2} = \frac{n_o}{n} \bar{Y}_{a2o} + \frac{n_d}{n} \bar{Y}_{a2d,k} - \bar{Y}_{r2}$$

where,

$$\bar{Y}_{a2d,k} = \frac{1}{n_d} \sum_{i \in \mathcal{D}} \tilde{Y}_{ai2,k} = \bar{Y}_{a2,o} + u_k + \left(\frac{r}{q} + b_k\right) (\bar{Y}_{a1d} - \bar{Y}_{a1,o}) + \bar{e}_k$$

and $\bar{e}_k = \frac{1}{n_d} \sum_{i \in \mathcal{D}} e_{i,k}$ and $\bar{Y}_{a1d} = \frac{1}{n_d} \sum_{i \in \mathcal{D}} Y_{ai1}$. Averaging the sample mean estimate for the k^{th} imputation, over the imputation set, K , gives Rubin's MI estimator as,

$$\hat{\theta}_{\text{DJ, MI}} = \frac{n_o}{n} \bar{Y}_{a2o} + \frac{n_d}{n} \left[\bar{Y}_{a2,o} + \bar{u} + \left(\frac{r}{q} + \bar{b}\right) (\bar{Y}_{a1d} - \bar{Y}_{a1,o}) + \bar{e} \right] - \bar{Y}_{r2},$$

where $\bar{u} = \frac{1}{K} \sum_{k=1}^K u_k$, $\bar{b} = \frac{1}{K} \sum_{k=1}^K b_k$ and $\bar{e} = \frac{1}{K} \sum_{k=1}^K \bar{e}_k$. Taking the necessary expectations reveals,

$$E(\hat{\theta}_{\text{DJ, MI}}) = \frac{n_o}{n} \mu_{a2} + \frac{n_d}{n} \mu_{a2} - \mu_{r2} = \mu_{a2} - \mu_{r2},$$

which as required, is unbiased for the treatment difference. For Rubin's MI variance estimator, under MAR, we require,

$$E(V_{\text{DJ, MI}}) = E(\hat{W}) + \left(1 + \frac{1}{K}\right) E(\hat{B}),$$

where, \hat{W} and, \hat{B} are defined as in (1.7) and (1.8). First we consider $E(\hat{W})$,

$$E(\hat{W}) = E(\hat{\sigma}_k^2) = E\left(\frac{\frac{1}{n-1} \sum_{i=1}^n (Y_{ri2} - \hat{\mu}_{r2})^2}{n}\right) + E\left(\frac{\frac{1}{n-1} \left[\sum_{i \in \mathcal{O}} (Y_{ai2} - \hat{\mu}_{a2,k})^2 + \sum_{i \in \mathcal{D}} (\hat{Y}_{ai2,k} - \hat{\mu}_{a2,k})^2 \right]}{n}\right).$$

Expanding these sums where required and taking the necessary expectations, (see Appendix B.4.1 for detailed calculations) for time 1 a baseline, where $\mu_{a1} = \mu_{r1}$, $\rho^2 = \sigma_{12}^2 / \sigma_{11}\sigma_{22}$ (the correlation between time 1 and time 2 squared) and $\sigma_{2.1} = \sigma_{22}(1 - \rho^2)$ gives,

$$E\left(\frac{1}{n-1} \sum_{i=1}^n (Y_{ri2} - \hat{\mu}_{r2})^2\right) = \sigma_{22},$$

$$E\left(\frac{1}{n-1} \left[\sum_{i \in \mathcal{O}} (Y_{ai2} - \hat{\mu}_{a2,k})^2 + \sum_{i \in \mathcal{D}} (\hat{Y}_{ai2,k} - \hat{\mu}_{a2,k})^2 \right]\right) = \sigma_{22} + \frac{2n_d\sigma_{22}(1 - \rho^2)}{(n_o - 1)(n - 1)}. \quad (3.4)$$

We now consider $(1 + \frac{1}{K}) E(\hat{B})$ where $E[\hat{B}] = \frac{1}{K-1} E\left[\sum_{k=1}^K (\hat{\theta}_{\text{DJ}, k} - \hat{\theta}_{\text{DJ}, \text{MI}})^2\right]$. Taking the necessary expectations (see Appendix B.4.1) reveals,

$$\left(1 + \frac{1}{K}\right) E(\hat{B}) = \left(1 + \frac{1}{K}\right) \sigma_{22}(1 - \rho^2) \frac{n_d}{n(n_o - 1)}. \quad (3.5)$$

Assuming that n is sufficiently large so that we may take $(n - 1)$ to be n and $(n_o - 1)$ to be n_o and where $\pi_d = n_d/n$, gives Rubin's variance estimator asymptotically as,

$$E[V_{\text{DJ}, \text{MI}}] = \frac{2\sigma_{22}}{n} + \frac{2\sigma_{22}(1 - \rho^2)\pi_d}{n^2(1 - \pi_d)} + \left(1 + \frac{1}{K}\right) \frac{\sigma_{22}(1 - \rho^2)\pi_d}{n(1 - \pi_d)}.$$

For infinite K this is equivalent to,

$$E[V_{\text{DJ}, \text{MI}}] = \sigma_{22} \left[\frac{1}{n} + \frac{1}{n_o} \right] - \frac{\sigma_{12}^2}{\sigma_{11}} \frac{\pi_d}{n(1 - \pi_d)} + \frac{2\sigma_{22}(1 - \rho^2)\pi_d}{n^2(1 - \pi_d)}. \quad (3.6)$$

Table B.1 summarises Rubin's derived variance under MAR alongside simulation results for various proportions of missing data in a setting inspired by the baseline and week 12 data in the asthma trial, introduced in Section 1.7.2. In all settings the derived results are within 2 MCSE's of the

simulation results.

This result provides us with key insight into the behaviour of Rubin's MI variance estimator. Under MAR we know valid inference is given from a full ML analysis of the observed data. The expectation of the variance of the treatment effect from a full ML analysis is (see Appendix B.2 for detailed calculations),

$$E[V_{\text{DJ, ML}}] = \sigma_{22} \left[\frac{1}{n} + \frac{1}{n_o} \right] - \frac{\sigma_{12}^2}{\sigma_{11}} \frac{\pi_d}{n(1-\pi_d)}. \quad (3.7)$$

We see Rubin's MI variance estimator (3.6) indeed approximates the variance estimate from a full likelihood analysis of the observed data (3.7), which accounts for the loss of information, giving a valid estimate of the variance in the observed incomplete data under MAR. The difference between (3.6) and (3.7) is $O(n^{-2})$, which is smaller in magnitude than $E[V_{\text{DJ, ML}}]$. We see the information in complete data relative to that in incomplete data as measured by Rubin's MI variance estimator under MAR (with $K \rightarrow \infty$) is,

$$\begin{aligned} \frac{I_{\text{DJ, full}}}{I_{\text{DJ, MI}}} &= 1 + \frac{(1-\rho^2)}{2} \left[\frac{n+1}{n} \pi_d + \frac{2n+1}{2n} \pi_d^2 + \frac{4n+1}{4n} \pi_d^3 + \frac{8n+1}{8n} \pi_d^4 \dots \right] \\ &\approx 1 + \frac{(1-\rho^2)}{2} [\pi_d + \pi_d^2 + \pi_d^3 + \pi_d^4 \dots]. \end{aligned} \quad (3.8)$$

This is asymptotically equivalent to the information in complete data relative to that in incomplete data from a full ML analysis, which is,

$$\frac{I_{\text{DJ, full}}}{I_{\text{DJ, ML}}} = 1 + \frac{(1-\rho^2)}{2} [\pi_d + \pi_d^2 + \pi_d^3 + \pi_d^4 \dots].$$

We have now derived the inflation factor which provides a measure of the loss of information due to missing data in the primary design analysis. The required information anchored variance estimator in sensitivity analysis is the design based variance estimator we would obtain for the treatment effect, were we able to observe the post-deviation data under reference based deviation, multiplied by this factor (3.8). We will now assess whether Rubin's variance estimator provides an acceptable estimate of the ideal information anchored variance estimate (3.2) in the reference based MI settings.

3.2 Rubin's variance estimator

We now analytically derive the expectation of Rubin's MI variance estimator in each of the de-facto settings (CR, J2R, CIR and LMCF) and compare this to the expectation of the information anchored variance estimate, which preserves the fraction of missing information in the primary

design based analysis.

Henceforth, in calculations for $E[V_{\text{DJ, MI}}]$, we take up to the $O(n^{-2})$ terms to simplify subsequent expressions, since this is equivalent to the MAR ML variance (3.7). Our inferences below do not change in doing so, since we would only have additional very small terms. Additionally we maintain our assumption of a known variance throughout, without any serious loss of generality [44].

3.2.1 Copy reference

Here we consider the case of CR behaviour post-deviation, and suppose that the partially observed patients in the active arm (i.e. those missing at time 2) have a time 2 mean equal to μ_{r2} .

Rubin's variance estimator under copy reference

As under MAR, an appropriate imputation distribution for the missing outcomes at time 2 is required, which also requires a suitable posterior for the included parameters. Under CR the data follows the distribution of the reference arm throughout the trial. For imputation we therefore require the conditional distribution of Y_{ri2} on Y_{ri1} .

To simplify subsequent calculations we define $Y_{ri1}^* = Y_{ri1} - \bar{Y}_{r1}$ where $\bar{Y}_{r1} = \frac{1}{n} \sum_{i=1}^n Y_{ri1}$. The observed data at time 2 in the reference arm is denoted by the vector \mathbf{Y}_{r2} where,

$$\mathbf{Y}_{r2} = (Y_{r12}, \dots, Y_{rn2})^T.$$

The data from the reference arm at time 1 is denoted by the vector,

$$\mathbf{Y}_{r1}^* = (Y_{r11}^*, \dots, Y_{rn1}^*)^T.$$

Under CR, our imputation model is formed from the regression of Y_{ri2} on Y_{ri1}^* for the n patients in the reference arm. Let $\mathbf{x}_i = (1, Y_{ri1}^*)^T$ then the conditional model of interest is,

$$Y_{ri2} = \mathbf{x}_i^T \beta + e_i, \quad e_i \sim N(0, \sigma_{2.1}) \quad i = 1, \dots, n,$$

where $\sigma_{2.1}$ is the residual variance from the regression of Y_{ri2} on Y_{ri1}^* . Re-expressed in matrix notation the conditional distribution of interest for imputation $\mathbf{Y}_{r2} | \mathbf{Y}_{r1}^*$ is,

$$(\mathbf{Y}_{r2} | \mathbf{Y}_{r1}^*) \sim N(\mathbf{X}\beta; \sigma_{2.1}),$$

where \mathbf{X} is the design matrix,

$$\mathbf{X} = \begin{bmatrix} 1 & Y_{r11}^* \\ 1 & \cdot \\ 1 & Y_{rn1}^* \end{bmatrix}.$$

The ordinary least square estimates, $\hat{\beta}$, are obtained by fitting the conditional model to the observed data and calculating $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}_{r2}$ which gives,

$$\hat{\beta} = \begin{pmatrix} \bar{Y}_{r2} \\ r/q \end{pmatrix},$$

where $\bar{Y}_{r2} = \frac{1}{n} \sum_{i=1}^n Y_{ri2}$, and under CR we have $r = \sum_{i=1}^n (Y_{ri1} - \bar{Y}_{r1}) (Y_{ri2} - \bar{Y}_{r2})$ and $q = \sum_{i=1}^n (Y_{ri1} - \bar{Y}_{r1})^2$, with associated variance $\sigma_{2.1} (\mathbf{X}^T \mathbf{X})^{-1}$. That is,

$$\text{Var}(\hat{\beta}) = \sigma_{2.1} \begin{pmatrix} n^{-1} & 0 \\ 0 & q^{-1} \end{pmatrix},$$

for $\sigma_{2.1} = \sigma_{22} - \sigma_{12}^2 / \sigma_{11}$. We assume the data is normal, the variance is known, improper priors for the missing observations and so the posterior of the model parameters is approximately normal and let the observed data dominate the prior. Then the CR imputation model for patient i and imputation k is,

$$\tilde{Y}_{ai2,k} = \bar{Y}_{r2} + u_k + (r/q + b_k) (Y_{ai1} - \bar{Y}_{r1}) + e_{i,k}, \text{ for } i \in \mathcal{D}, k = 1, \dots, K, \quad (3.9)$$

where,

$$\begin{aligned} u_k &\sim N(0, n^{-1} \sigma_{2.1}) \\ b_k &\sim N(0, q^{-1} \sigma_{2.1}) \\ e_{i,k} &\sim N(0, \sigma_{2.1}). \end{aligned}$$

Using the above imputation model, the treatment effect for the k^{th} imputation over the missing

cases is obtained as,

$$\hat{\theta}_{\text{CR}, k} = \hat{\mu}_{a2,k} - \hat{\mu}_{r2} = \frac{n_o}{n} \bar{Y}_{a2o} + \frac{n_d}{n} \bar{Y}_{a2d,k} - \bar{Y}_{r2},$$

where,

$$\bar{Y}_{a2d,k} = \frac{1}{n_d} \sum_{i \in \mathcal{D}} \tilde{Y}_{ai2,k} = \bar{Y}_{r2} + u_k + \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{r1}) + \bar{e}_k,$$

and $\bar{e}_k = \frac{1}{n_d} \sum_{i \in \mathcal{D}} e_{i,k}$ and $\bar{Y}_{a1d} = \frac{1}{n_d} \sum_{i \in \mathcal{D}} Y_{ai1}$. Averaging the sample mean estimate for the k^{th} imputation, over the imputation set, K , gives Rubin's MI estimator as,

$$\hat{\theta}_{\text{CR}, \text{MI}} = \frac{n_o}{n} \bar{Y}_{a2o} + \frac{n_d}{n} \left[\bar{Y}_{r2} + \bar{u} + \left(\frac{r}{q} + \bar{b} \right) (\bar{Y}_{a1d} - \bar{Y}_{r1}) + \bar{e} \right] - \bar{Y}_{r2},$$

where $\bar{u} = \frac{1}{K} \sum_{k=1}^K u_k$, $\bar{b} = \frac{1}{K} \sum_{k=1}^K b_k$ and $\bar{e} = \frac{1}{K} \sum_{k=1}^K \bar{e}_k$. Taking the necessary expectations reveals,

$$E(\hat{\theta}_{\text{CR}, \text{MI}}) = \frac{n_o}{n} (\mu_{a2} - \mu_{r2}),$$

which as required, is unbiased for the treatment effect under CR. This is unsurprising given the results of the simulation study in Section 2.3. We explicitly see how the imputation only incorporates additive terms which have an expectation of zero, thus how taking the average treatment effect over the imputation set provides an unbiased estimator. For Rubin's MI variance estimator, under CR, we require,

$$E(V_{\text{CR}, \text{MI}}) = E(\hat{W}) + \left(1 + \frac{1}{K} \right) E(\hat{B}),$$

where, \hat{W} and, \hat{B} are defined as in (1.7) and (1.8). First we consider $E(\hat{W})$,

$$E(\hat{W}) = E(\hat{\sigma}_k^2) = E\left(\frac{\frac{1}{n-1} \sum_{i=1}^n (Y_{ri2} - \hat{\mu}_{r2})^2}{n}\right) + E\left(\frac{\frac{1}{n-1} \left[\sum_{i \in \mathcal{O}} (Y_{ai2} - \hat{\mu}_{a2,k})^2 + \sum_{i \in \mathcal{D}} (\hat{Y}_{ai2,k} - \hat{\mu}_{a2,k})^2 \right]}{n}\right). \quad (3.10)$$

We define the first component on the RHS of (3.10) as $W1$ and the second component on the RHS of (3.10) as $W2$. Expanding these sums where required and taking the necessary expectations, (see Appendix B.4.2 for detailed calculations) for time 1 a baseline, where $\mu_{a1} = \mu_{r1}$, $\pi_d = n_d/n$ and $\Delta = \mu_{a2} - \mu_{r2}$ gives,

$$E(W1) = \frac{\sigma_{22}}{n},$$

$$E(W2) = \frac{\sigma_{22}}{n} + \frac{\Delta^2 \pi_d n_o}{n(n-1)} + \frac{2\sigma_{22}(1-\rho^2)\pi_d}{n(n-1)} \left[2 - \pi_d \left(1 + \frac{1}{n} \right) \right]. \quad (3.11)$$

We now consider $(1 + \frac{1}{K}) E(\hat{B})$. Taking the necessary expectations reveals,

$$\left(1 + \frac{1}{K} \right) E(\hat{B}) = \left(1 + \frac{1}{K} \right) \sigma_{22} (1 - \rho^2) \pi_d \frac{(\pi_d + 1)}{(n-1)}. \quad (3.12)$$

Assuming that n is sufficiently large so that we may take $(n-1)$ to be n gives Rubin's estimate asymptotically as,

$$E[V_{\text{CR, MI}}] = \frac{2\sigma_{22}}{n} + \frac{\Delta^2 \pi_d (1 - \pi_d)}{n} + \frac{2\sigma_{22}(1-\rho^2)\pi_d}{n^2} \left[2 - \pi_d \left(1 + \frac{1}{n} \right) \right] + \left(1 + \frac{1}{K} \right) \sigma_{22} (1 - \rho^2) \pi_d \frac{(\pi_d + 1)}{n}.$$

For infinite K this is equivalent to,

$$E[V_{\text{CR, MI}}] = \frac{2\sigma_{22}}{n} + \frac{\Delta^2 \pi_d (1 - \pi_d)}{n} + \frac{2\sigma_{22}(1-\rho^2)\pi_d}{n^2} \left[2 - \pi_d \left(1 + \frac{1}{n} \right) \right] + \sigma_{22} (1 - \rho^2) \pi_d \frac{(\pi_d + 1)}{n}. \quad (3.13)$$

Table B.3 summarises Rubin's derived variance under CR alongside simulation results. In all

missingness settings the derivation results are within 2 MCSE's of the results.

Information anchoring under copy reference

Consider the expectation of the ideal information anchored variance estimate that preserves the fraction of missing information in the primary design based analysis under CR. This is calculated as (3.8)×(3.1). For $\Delta = (\mu_{a2} - \mu_{r2})$ up to the order π_d^4/n terms this is,

$$E[V_{\text{anchored}}] = \frac{2\sigma_{22}}{n} + \frac{\Delta^2 \pi_d (1 - \pi_d)}{n} + \frac{\sigma_{22} (1 - \rho^2) \pi_d}{n} + \left[\frac{\sigma_{22}}{n} + \frac{\Delta^2}{2n} \right] (1 - \rho^2) \pi_d^2 + \frac{\sigma_{22} (1 - \rho^2) \pi_d^3}{n} + \frac{\sigma_{22} (1 - \rho^2) \pi_d^4}{n}. \quad (3.14)$$

Now compare this to Rubin's CR MI variance estimator where $K \rightarrow \infty$ (3.13), which can be re-expressed as,

$$E[V_{\text{CR, MI}}] = \frac{2\sigma_{22}}{n} + \frac{\Delta^2 \pi_d (1 - \pi_d)}{n} + \frac{\sigma_{22} (1 - \rho^2) \pi_d}{n} + \frac{\sigma_{22} (1 - \rho^2) \pi_d^2}{n} + \frac{2\sigma_{22} (1 - \rho^2) \pi_d}{n^2} \left[2 - \pi_d \left(1 + \frac{1}{n} \right) \right].$$

The difference (3.13) - (3.14) is computed as,

$$E[V_{\text{CR, MI}}] - E[V_{\text{anchored}}] = \frac{2\sigma_{22} (1 - \rho^2) \pi_d}{n^2} \left[2 - \pi_d \left(1 + \frac{1}{n} \right) \right] - \frac{\Delta^2 (1 - \rho^2) \pi_d^2}{2n} - \frac{\sigma_{22} (1 - \rho^2) \pi_d^3}{n} - \frac{\sigma_{22} (1 - \rho^2) \pi_d^4}{n}. \quad (3.15)$$

The largest order terms in the difference between Rubin's MI variance and the ideal variance under CR are,

$$-\frac{\Delta^2 (1 - \rho^2) \pi_d^2}{2n} - \frac{\sigma_{22} (1 - \rho^2) \pi_d^3}{n} - \frac{\sigma_{22} (1 - \rho^2) \pi_d^4}{n}. \quad (3.16)$$

We see there is excellent agreement. The difference depends on the size of the trial (n), the amount of missingness in the active arm (n_d), the correlation between the time 1 and time 2 measurement (ρ^2), the variance of the data at time 2 (σ_{22}), as well as the mean difference between the active and reference arm at time 2 (Δ). Rubin's variance estimator will better approximate the information anchored variance in larger trials with smaller amounts of missingness, higher correlation between the time 1 and time 2 measurements and smaller differences between the arms at time 2. However,

we see from its composition that the difference will still be small in settings with greater amounts of missingness, smaller correlations between time 1 and time 2 measurements and greater differences between the arms at time 2. For larger n the difference quickly approaches zero, at a faster rate than that of the required information anchored variance. The information anchored variance (3.14) is dominated by the $2\sigma_{22}/n$ term, which is $O(n^{-1})$.

The ratio of the difference between Rubin's variance estimator and the information anchored variance (3.16) to the information anchored variance (3.14) is,

$$\frac{(1 - \rho^2) \pi_d^2 [-\Delta^2 - 2\sigma_{22} (\pi_d + \pi_d^2)]}{4\sigma_{22} + \pi_d [\Delta^2 (2(1 - \pi_d) + (1 - \rho^2) \pi_d) + 2\sigma_{22} (1 - \rho^2) (1 + \pi_d + \pi_d^2 + \pi_d^3)]}.$$

We recall here the simple sample size formula for detecting a difference in means between two independent groups [58],

$$n = \frac{2\sigma_{22} (Z_\beta + Z_{\frac{\alpha}{2}})^2}{(\mu_1 - \mu_2)^2},$$

where n is the sample size required in each group, Z_β is the one-sided percentage point of the normal distribution corresponding to the required power, $Z_{\frac{\alpha}{2}}$ is the percentage point of the normal distribution corresponding to the two-sided significance level, σ_{22} is the variance of the outcome variable and $\mu_1 - \mu_2$ represents the difference in means to detect. In a typical RCT powered at 80% ($Z_\beta = 0.84$) with 5% ($Z_{\frac{\alpha}{2}} = 1.96$) statistical significance therefore,

$$\Delta^2 \approx (\mu_1 - \mu_2)^2 = \frac{15.68\sigma_{22}}{n}. \quad (3.17)$$

Thus the first term in (3.16) is very small and $O(n^{-2})$. The second two components in (3.16) are $O(\pi_d^3 n^{-1})$ and $O(\pi_d^4 n^{-1})$. Since $\pi_d < 1$ in sensitivity analysis, these terms will also be tiny in practice. For realistic proportions of missing data i.e. $\pi_d \leq 0.5$, these two terms are at most $O((8n)^{-1})$ and $O((16n)^{-1})$ respectively. Thus all these terms are very small relative to the information anchored variance we desire.

We can use (3.17) to express the difference between Rubin's variance estimator and the information anchored variance and the ratio of this difference to the information anchored variance, in terms of the variance σ_{22} . Table 3.1 summarises the difference between Rubin's CR MI variance estimator and the information anchored variance estimate for various sized trials and proportions of missing data with $\rho = 0.5$, in multiples σ_{22} .

We clearly see how in practice the agreement between Rubin's variance estimator and the information anchored variance will be excellent, even for smaller sample sizes. With small proportions of missing data we see Rubin's variance estimator will be marginally overestimating the information anchored variance. This is because the positive terms in (3.15) will dominate the difference. With

larger proportions of missing data Rubin’s variance estimator will be marginally underestimating the information anchored variance as the negative terms in (3.15) become larger. However in all cases the difference between Rubin’s variance estimator and the information anchored variance is negligible in size, in comparison to the information anchored variance. The ratio of the difference to the information anchored variance is tiny in all cases. Rubin’s variance estimator provides an excellent approximation for the ideal information anchored variance.

n	Proportion of Missingness (active)	Information anchored variance (in multiples of σ_{22})	Difference (in multiples of σ_{22})	Difference/Information anchored variance (in multiples of σ_{22})
20	0.1	0.107841	0.000522	0.004840
	0.2	0.116220	0.000395	0.003399
	0.3	0.125496	-0.000744	-0.005928
	0.4	0.136120	-0.003342	-0.024552
	0.5	0.148631	-0.007941	-0.053428
50	0.1	0.042255	0.000074	0.001751
	0.2	0.044842	-0.000023	-0.000513
	0.3	0.047905	-0.000433	-0.009039
	0.4	0.051626	-0.001338	-0.025917
	0.5	0.056219	-0.002954	-0.052545
100	0.1	0.020980	0.000014	0.000667
	0.2	0.022146	-0.000042	-0.001897
	0.3	0.023571	-0.000240	-0.010182
	0.4	0.025342	-0.000670	-0.026438
	0.5	0.027570	-0.001441	-0.052267
250	0.1	0.008357	0.000000	0.000000
	0.2	0.008793	-0.000024	-0.002729
	0.3	0.009337	-0.000102	-0.010924
	0.4	0.010024	-0.000269	-0.026836
	0.5	0.010899	-0.000568	-0.052115

Table 3.1: Difference between Rubin’s CR MI variance estimator and the information anchored variance with $\rho = 0.5$ using (3.17)

3.2.2 Jump to reference

We now derive the expectation of Rubin’s variance estimator in the J2R scenario and compare this with the information anchored variance.

Rubin’s variance estimator under jump to reference

The data distribution for active deviators under J2R is formed using: the mean from the active arm at time 1, the mean from the reference arm at time 2, variance from the active arm at time 1 and variance from the reference arm for time 2 and the conditional components (time 2 given time 1). Appendix B.3 details the steps taken to form the conditional model this implies for time 2 given time 1. Assuming the data is normal, variance is known and improper priors for

the missing observations, then the posterior of the model parameters is approximately normal, with prior dominated by the observed data. Then the J2R imputation model for patient i and imputation k is,

$$\begin{aligned} \tilde{Y}_{ai2,k} = \bar{Y}_{r2} + u_k + (r/q + b_k) (Y_{ai1} - \bar{Y}_{a1}) + (r/q + b_k) (r_k - a_k) + e_{i,k}, \\ \text{for } i \in \mathcal{D}, k = 1, \dots, K, \end{aligned} \quad (3.18)$$

where,

$$\begin{aligned} u_k &\sim N(0, n^{-1}\sigma_{2.1}) \\ b_k &\sim N(0, q^{-1}\sigma_{2.1}) \\ r_k &\sim N(0, \frac{\sigma_{11}}{n}) \\ a_k &\sim N(0, \frac{\sigma_{11}}{n}) \\ e_{i,k} &\sim N(0, \sigma_{2.1}), \end{aligned}$$

and under J2R $r = \sum_{i=1}^n (Y_{ri1} - \bar{Y}_{r1}) (Y_{ri2} - \bar{Y}_{r2})$ and $q = \sum_{i=1}^n (Y_{ri1} - \bar{Y}_{r1})^2$. Using the above imputation model we obtain the treatment effect for the k^{th} imputation over the missing cases as,

$$\hat{\theta}_{\text{J2R}, k} = \hat{\mu}_{a2,k} - \hat{\mu}_{r2} = \frac{n_o}{n} \bar{Y}_{a2o} + \frac{n_d}{n} \bar{Y}_{a2d,k} - \bar{Y}_{r2},$$

where,

$$\bar{Y}_{a2d,k} = \frac{1}{n_d} \sum_{i \in \mathcal{D}} \tilde{Y}_{ai2,k} = \bar{Y}_{r2} + u_k + \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{a1}) + (r/q + b_k) (r_k - a_k) + \bar{e}_k,$$

and $\bar{e}_k = \frac{1}{n_d} \sum_{i \in \mathcal{D}} e_{i,k}$ and $\bar{Y}_{a1d} = \frac{1}{n_d} \sum_{i \in \mathcal{D}} Y_{ai1}$. Averaging the sample mean estimate for the k^{th} imputation, over the imputation set, K , gives Rubin's MI estimator as,

$$\begin{aligned} \hat{\theta}_{\text{J2R}, \text{MI}} = \frac{n_o}{n} \bar{Y}_{a2o} + \frac{n_d}{n} \left[\bar{Y}_{r2} + \bar{u} + \left(\frac{r}{q} + \bar{b} \right) (\bar{Y}_{a1d} - \bar{Y}_{a1}) + \frac{1}{K} \sum_{k=1}^K (r/q + b_k) (r_k - a_k) + \bar{e} \right] \\ - \bar{Y}_{r2} \end{aligned}$$

where $\bar{u} = \frac{1}{K} \sum_{k=1}^K u_k$, $\bar{b} = \frac{1}{K} \sum_{k=1}^K b_k$ and $\bar{e} = \frac{1}{K} \sum_{k=1}^K \bar{e}_k$. Taking the necessary expectations reveals,

$$E(\hat{\theta}_{\text{CR, MI}}) = \frac{n_o}{n} (\mu_{a2} - \mu_{r2}),$$

which as required is unbiased for the J2R treatment effect, which with baseline and a single follow-up is the same as the CR treatment estimate. For Rubin's MI variance estimator, under J2R, we require,

$$E(V_{\text{J2R, MI}}) = E(\hat{W}) + \left(1 + \frac{1}{K}\right) E(\hat{B}),$$

where, \hat{W} and, \hat{B} are defined as in (1.7) and (1.8). First we consider $E(\hat{W})$,

$$\begin{aligned} E(\hat{W}) = E(\hat{\sigma}_k^2) = & E\left(\frac{\frac{1}{n-1} \sum_{i=1}^n (Y_{ri2} - \hat{\mu}_{r2})^2}{n}\right) \\ & + E\left(\frac{\frac{1}{n-1} \left[\sum_{i \in \mathcal{O}} (Y_{ai2} - \hat{\mu}_{a2,k})^2 + \sum_{i \in \mathcal{D}} (\hat{Y}_{ai2,k} - \hat{\mu}_{a2,k})^2\right]}{n}\right). \end{aligned} \quad (3.19)$$

We define the first component on the RHS of (3.19) as $W1$ and the second component on the RHS of (3.19) as $W2$. Expanding these sums where required and taking the necessary expectations, (see Appendix B.4.3 for detailed calculations) for time 1 a baseline, where $\mu_{a1} = \mu_{r1}$, $\pi_d = n_d/n$ and $\Delta = \mu_{a2} - \mu_{r2}$ gives,

$$\begin{aligned} E(W1) &= \frac{\sigma_{22}}{n}, \\ E(W2) &= \frac{\sigma_{22}}{n} + \frac{\Delta^2 \pi_d n_o}{n(n-1)} + \frac{\sigma_{22} \pi_d (2 - \pi_d)}{n(n-1)} + \frac{\sigma_{12}^2 \pi_d (2 - 3\pi_d)}{\sigma_{11} n(n-1)} \\ &+ \frac{\sigma_{22} (1 - \rho^2) \pi_d}{n(n-1)} \left[\frac{2n}{(n-1)} - \frac{2\pi_d}{(n-1)} - \pi_d \right]. \end{aligned} \quad (3.20)$$

We now consider $(1 + \frac{1}{K}) E(\hat{B})$. Taking the necessary expectations reveals,

$$\left(1 + \frac{1}{K}\right) E(\hat{B}) = \left(1 + \frac{1}{K}\right) \pi_d^2 \left[\frac{\sigma_{22} (1 - \rho^2)}{(n-1)} \left(\frac{2}{n} + 1 + \frac{1}{\pi_d}\right) + \frac{2 \sigma_{12}^2}{n \sigma_{11}} \right]. \quad (3.21)$$

Assuming that n is sufficiently large so that we may take $(n-1)$ to be n gives Rubin's variance

estimator asymptotically as,

$$\begin{aligned}
E[V_{\text{J2R, MI}}] &= \frac{2\sigma_{22}}{n} + \frac{\Delta^2 \pi_d (1 - \pi_d)}{n} + \frac{\sigma_{22} \pi_d (2 - \pi_d)}{n^2} + \frac{\sigma_{12}^2 \pi_d (2 - 3\pi_d)}{\sigma_{11} n^2} \\
&\quad + \frac{\sigma_{22} (1 - \rho^2) \pi_d}{n^2} \left[2 - \frac{2\pi_d}{n} - \pi_d \right] \\
&\quad + \left(1 + \frac{1}{K} \right) \frac{\pi_d^2}{n} \left[\sigma_{22} (1 - \rho^2) \left(\frac{2}{n} + 1 + \frac{1}{\pi_d} \right) + 2 \frac{\sigma_{12}^2}{\sigma_{11}} \right].
\end{aligned}$$

For infinite K this is,

$$\begin{aligned}
E[V_{\text{J2R, MI}}] &= \frac{2\sigma_{22}}{n} + \frac{\Delta^2 \pi_d (1 - \pi_d)}{n} + \frac{\sigma_{22} (1 - \rho^2) \pi_d}{n} + \frac{\sigma_{22} (1 - \rho^2) \pi_d^2}{n} + 2 \frac{\sigma_{12}^2 \pi_d^2}{\sigma_{11} n} \\
&\quad + \frac{\sigma_{12}^2 \pi_d (2 - 3\pi_d)}{\sigma_{11} n^2} + \frac{\sigma_{22} (1 - \rho^2) \pi_d}{n^2} \left[2 + \pi_d \left(1 - \frac{2}{n} \right) \right] + \frac{\sigma_{22} \pi_d (2 - \pi_d)}{n^2}. \quad (3.22)
\end{aligned}$$

Table B.5 shows the derived results for Rubin's variance estimator under J2R are within 2 MCSE's of simulation results.

Information anchoring under jump to reference

Consider the expectation of the ideal information anchored variance estimate under J2R. For $\Delta = (\mu_{a2} - \mu_{r2})$ up to the order π_d^4/n terms this is,

$$\begin{aligned}
E[V_{\text{anchored}}] &= \frac{2\sigma_{22}}{n} + \frac{\Delta^2 \pi_d (1 - \pi_d)}{n} + \frac{\sigma_{22} (1 - \rho^2) \pi_d}{n} + \left[\frac{\sigma_{22}}{n} + \frac{\Delta^2}{2n} \right] (1 - \rho^2) \pi_d^2 \\
&\quad + \frac{\sigma_{22} (1 - \rho^2) \pi_d^3}{n} + \frac{\sigma_{22} (1 - \rho^2) \pi_d^4}{n}. \quad (3.23)
\end{aligned}$$

We now compare this to Rubin's J2R MI variance estimator where $K \rightarrow \infty$ (3.22). The difference (3.22) - (3.23) is,

$$\begin{aligned}
E[V_{\text{J2R, MI}}] - E[V_{\text{anchored}}] &= 2 \frac{\sigma_{12}^2 \pi_d^2}{\sigma_{11} n} + \frac{\sigma_{12}^2 \pi_d (2 - 3\pi_d)}{\sigma_{11} n^2} \\
&\quad + \frac{\sigma_{22} (1 - \rho^2) \pi_d}{n^2} \left[2 + \pi_d \left(1 - \frac{2}{n} \right) \right] + \frac{\sigma_{22} \pi_d (2 - \pi_d)}{n^2} \\
&\quad - \frac{\Delta^2 (1 - \rho^2) \pi_d^2}{2n} - \frac{\sigma_{22} (1 - \rho^2) \pi_d^3}{n} - \frac{\sigma_{22} (1 - \rho^2) \pi_d^4}{n}.
\end{aligned}$$

The largest order terms in the difference between Rubin's MI variance and the ideal variance under J2R are,

$$2 \frac{\sigma_{12}^2 \pi_d^2}{\sigma_{11} n} - \frac{\Delta^2 (1 - \rho^2) \pi_d^2}{2n} - \frac{\sigma_{22} (1 - \rho^2) \pi_d^3}{n} - \frac{\sigma_{22} (1 - \rho^2) \pi_d^4}{n}. \quad (3.24)$$

We see the difference depends on the same elements as identified for the CR scenario: trial size, proportion of missing data in the active arm, correlation between time 1 and time 2 measurements and the difference in means between the arms at time 1 and time 2. With smaller amounts of missingness, higher correlations and smaller effect sizes Rubin's variance estimator will preserve the fraction of missing information in the primary design based analysis more closely. However following the inferences in Subsection 3.2.1 we highlight how small the terms in (3.24) actually are. The second term in (3.24) will be $O(n^{-2})$ and the third and fourth terms will respectively be $O(\pi_d^3 n^{-1})$ and $O(\pi_d^4 n^{-1})$. As previously discussed these terms will be negligible in practice. The first term in (3.24) is $O(\pi_d^2 n^{-1})$. This will also be of a very small size in practice in comparison to the information anchored variance we desire —(3.23) which is $O(n^{-1})$ — since $\pi_d < 1$. The ratio of the difference between Rubin's variance estimator and the information anchored variance (3.24) to the information anchored variance (3.23) is,

$$\frac{(1 - \rho^2) \pi_d^2 [-\Delta^2 - 2\sigma_{22} (\pi_d + \pi_d^2)] + 4\sigma_{12}^2 \sigma_{11}^{-1} \pi_d^2}{4\sigma_{22} + \pi_d [\Delta^2 (2(1 - \pi_d) + (1 - \rho^2) \pi_d) + 2\sigma_{22} (1 - \rho^2) (1 + \pi_d + \pi_d^2 + \pi_d^3)]}.$$

Thus we infer that in practice agreement will be excellent. The difference is of no practical importance. Rubin's variance estimator also approximates the ideal information anchored variance under J2R.

3.2.3 Copy increments in reference

We now calculate the expectation of Rubin's variance estimator in the CIR scenario and compare this to the information anchored variance estimate.

Rubin's variance estimator under copy increments in reference

The data distribution under CIR is formed using: the mean from the active arm at time 1, the mean in the active arm at time 1 plus the difference in the means at time 2 and time 1 in the reference arm for time 2, variance from the active arm at time 1 and variance from the reference arm for time 2 and the conditional components. Appendix B.3 details the steps taken to form the conditional model this implies for time 2 given time 1. Again, assuming the data is normal, variance is known and improper priors for the missing observations, then the posterior of the model parameters is approximately normal, with prior dominated by the observed data. Then the CIR imputation model for patient i and imputation k is,

$$\begin{aligned}\tilde{Y}_{ai2,k} &= \bar{Y}_{a1} + \bar{Y}_{r2} - \bar{Y}_{r1} + u_k + (r/q + b_k)(Y_{ai1} - \bar{Y}_{a1}) + (r/q + b_k)(r_k - a_k) + a_k - r_k + e_{i,k}, \\ &\text{for } i \in \mathcal{D}, k = 1, \dots, K,\end{aligned}\tag{3.25}$$

where,

$$\begin{aligned}u_k &\sim N(0, n^{-1}\sigma_{2.1}) \\ b_k &\sim N(0, q^{-1}\sigma_{2.1}) \\ r_k &\sim N(0, \frac{\sigma_{11}}{n}) \\ a_k &\sim N(0, \frac{\sigma_{11}}{n}) \\ e_{i,k} &\sim N(0, \sigma_{2.1}),\end{aligned}$$

and for CIR we have $r = \sum_{i=1}^n (Y_{ri1} - \bar{Y}_{r1})(Y_{ri2} - \bar{Y}_{r2})$ and $q = \sum_{i=1}^n (Y_{ri1} - \bar{Y}_{r1})^2$. Using the above imputation model we obtain the treatment effect for the k^{th} imputation over the missing cases as,

$$\hat{\theta}_{\text{CIR}, k} = \hat{\mu}_{a2,k} - \hat{\mu}_{r2} = \frac{n_o}{n} \bar{Y}_{a2o} + \frac{n_d}{n} \bar{Y}_{a2d,k} - \bar{Y}_{r2},$$

where,

$$\begin{aligned}\bar{Y}_{a2d,k} &= \frac{1}{n_d} \sum_{i \in \mathcal{D}} \tilde{Y}_{ai2,k} = \bar{Y}_{a1} + \bar{Y}_{r2} - \bar{Y}_{r1} + u_k + \left(\frac{r}{q} + b_k\right) (\bar{Y}_{a1d} - \bar{Y}_{a1}) + (r/q + b_k)(r_k - a_k) \\ &\quad + a_k - r_k + \bar{e}_k,\end{aligned}$$

and $\bar{e}_k = \frac{1}{n_d} \sum_{i \in \mathcal{D}} e_{i,k}$ and $\bar{Y}_{a1d} = \frac{1}{n_d} \sum_{i \in \mathcal{D}} Y_{ai1}$. Averaging the sample mean estimate for the k^{th} imputation, over the imputation set, K , gives Rubin's MI estimator as,

$$\begin{aligned}\hat{\theta}_{\text{CIR, MI}} &= \frac{n_d}{n} \left[\bar{Y}_{r2} + \bar{u} + \left(\frac{r}{q} + \bar{b}\right) (\bar{Y}_{a1d} - \bar{Y}_{a1}) + \frac{1}{k} \sum_{k=1}^K (r/q + b_k)(r_k - a_k) + \bar{a} - \bar{r} + \bar{e} \right] \\ &\quad + \frac{n_o}{n} \bar{Y}_{a2o} - \bar{Y}_{r2},\end{aligned}$$

where $\bar{u} = \frac{1}{K} \sum_{k=1}^K u_k$, $\bar{b} = \frac{1}{K} \sum_{k=1}^K b_k$, $\bar{a} = \frac{1}{K} \sum_{k=1}^K a_k$, $\bar{r} = \frac{1}{K} \sum_{k=1}^K r_k$ and $\bar{e} = \frac{1}{K} \sum_{k=1}^K \bar{e}_k$.

Taking the necessary expectations reveals,

$$E(\hat{\theta}_{\text{CIR, MI}}) = \frac{n_o}{n} (\mu_{a2} - \mu_{r2}),$$

which is unbiased for the treatment estimate under CIR. We note this is also equivalent to the CR/J2R treatment estimate in the baseline and single follow-up setting. For Rubin's MI variance estimator we first require $E(\hat{W})$ under CIR,

$$E(\hat{W}) = E(\hat{\sigma}_k^2) = E\left(\frac{\frac{1}{n-1} \sum_{i=1}^n (Y_{ri2} - \hat{\mu}_{r2})^2}{n}\right) + E\left(\frac{\frac{1}{n-1} \left[\sum_{i \in \mathcal{O}} (Y_{ai2} - \hat{\mu}_{a2,k})^2 + \sum_{i \in \mathcal{D}} (\hat{Y}_{ai2,k} - \hat{\mu}_{a2,k})^2 \right]}{n}\right). \quad (3.26)$$

We again define the first component on the RHS of (3.26) as $W1$ and the second component on the RHS of (3.26) as $W2$. For time 1 a baseline where, $\mu_{a1} = \mu_{r1}$, $\pi_d = n_d/n$ and $\Delta = (\mu_{a2} - \mu_{r2})$ we obtain,

$$E(W1) = \frac{\sigma_{22}}{n},$$

$$E(W2) = \frac{\sigma_{22}}{n} + \frac{\Delta^2 \pi_d n_o}{n(n-1)} + \frac{\sigma_{22} \pi_d (2 - \pi_d)}{n(n-1)} + \frac{\sigma_{12}^2 \pi_d (2 - 3\pi_d)}{\sigma_{11} n(n-1)} + \frac{\sigma_{22} (1 - \rho^2) \pi_d}{n(n-1)} \left[\frac{2n}{(n-1)} - \frac{2\pi_d}{(n-1)} - \pi_d \right] + \frac{4\pi_d (1 - \pi_d)}{n(n-1)} (\sigma_{11} - 2\sigma_{12}). \quad (3.27)$$

We now consider $(1 + \frac{1}{K}) E(\hat{B})$. Taking the required expectations reveals,

$$\left(1 + \frac{1}{K}\right) E(\hat{B}) = \left(1 + \frac{1}{K}\right) \pi_d^2 \left[\frac{\sigma_{22} (1 - \rho^2)}{(n-1)} \left(\frac{2}{n} + 1 + \frac{1}{\pi_d}\right) + \frac{2}{n} \left(\frac{\sigma_{12}^2}{\sigma_{11}} + \sigma_{11} - 2\sigma_{12}\right) \right]. \quad (3.28)$$

Assuming that n is sufficiently large so that we may take $(n-1)$ to be n gives Rubin's variance estimator asymptotically as,

$$\begin{aligned}
E[V_{\text{CIR, MI}}] &= \frac{2\sigma_{22}}{n} + \frac{\Delta^2 \pi_d (1 - \pi_d)}{n} + \frac{\sigma_{22} \pi_d (2 - \pi_d)}{n^2} + \frac{\sigma_{12}^2 \pi_d (2 - 3\pi_d)}{\sigma_{11} n^2} \\
&\quad + \frac{\sigma_{22} (1 - \rho^2) \pi_d}{n^2} \left[2 - \frac{2\pi_d}{n} - \pi_d \right] + \frac{4\pi_d (1 - \pi_d)}{n^2} (\sigma_{11} - 2\sigma_{12}) \\
&\quad + \left(1 + \frac{1}{K} \right) \frac{\pi_d^2}{n} \left[\sigma_{22} (1 - \rho^2) \left(\frac{2}{n} + 1 + \frac{1}{\pi_d} \right) + 2 \left(\frac{\sigma_{12}^2}{\sigma_{11}} + \sigma_{11} - 2\sigma_{12} \right) \right].
\end{aligned}$$

For infinite K this is,

$$\begin{aligned}
E[V_{\text{CIR, MI}}] &= \frac{2\sigma_{22}}{n} + \frac{\Delta^2 \pi_d (1 - \pi_d)}{n} + \frac{\sigma_{22} (1 - \rho^2) \pi_d}{n} + \frac{\sigma_{22} (1 - \rho^2) \pi_d^2}{n} + 2 \frac{\sigma_{12}^2 \pi_d^2}{\sigma_{11} n} \\
&\quad + \frac{\sigma_{12}^2 \pi_d (2 - 3\pi_d)}{\sigma_{11} n^2} + \frac{\sigma_{22} (1 - \rho^2) \pi_d}{n^2} \left[2 + \pi_d \left(1 - \frac{2}{n} \right) \right] + \frac{\sigma_{22} \pi_d (2 - \pi_d)}{n^2} \\
&\quad + 2\pi_d (\sigma_{11} - 2\sigma_{12}) \left[\frac{(n - 2) \pi_d + 2}{n^2} \right]. \tag{3.29}
\end{aligned}$$

Table B.7 summarises Rubin's derived variance under CIR alongside simulation results. In all settings the derivation results are within 2 MCSE's of the simulation results.

Information anchoring under copy increments in reference

The expectation of the ideal information anchored variance estimate under CIR, for $\Delta = (\mu_{a2} - \mu_{r2})$ up to the order π_d^4/n terms is,

$$\begin{aligned}
E[V_{\text{anchored}}] &= \frac{2\sigma_{22}}{n} + \frac{\Delta^2 \pi_d (1 - \pi_d)}{n} + \frac{\sigma_{22} (1 - \rho^2) \pi_d}{n} + \left[\frac{\sigma_{22}}{n} + \frac{\Delta^2}{2n} \right] (1 - \rho^2) \pi_d^2 \\
&\quad + \frac{\sigma_{22} (1 - \rho^2) \pi_d^3}{n} + \frac{\sigma_{22} (1 - \rho^2) \pi_d^4}{n}. \tag{3.30}
\end{aligned}$$

The difference between Rubin's CIR MI variance estimator where $K \rightarrow \infty$ (3.29) and (3.30) is,

$$\begin{aligned}
E[V_{\text{CIR, MI}}] - E[V_{\text{anchored}}] &= 2 \frac{\sigma_{12}^2 \pi_d^2}{\sigma_{11} n} + \frac{\sigma_{12}^2 \pi_d (2 - 3\pi_d)}{\sigma_{11} n^2} + \frac{\sigma_{22} \pi_d (2 - \pi_d)}{n^2} \\
&\quad + \frac{\sigma_{22} (1 - \rho^2) \pi_d}{n^2} \left[2 + \pi_d \left(1 - \frac{2}{n} \right) \right] \\
&\quad + 2\pi_d (\sigma_{11} - 2\sigma_{12}) \left[\frac{(n - 2) \pi_d + 2}{n^2} \right] \\
&\quad - \frac{\Delta^2 (1 - \rho^2) \pi_d^2}{2n} - \frac{\sigma_{22} (1 - \rho^2) \pi_d^3}{n} - \frac{\sigma_{22} (1 - \rho^2) \pi_d^4}{n}.
\end{aligned}$$

The difference between Rubin's MI variance and the ideal variance under CIR is dominated by the following terms which have the largest magnitude,

$$2 \frac{\sigma_{12}^2 \pi_d^2}{\sigma_{11} n} + 2(\sigma_{11} - 2\sigma_{12}) \frac{\pi_d^2}{n} - \frac{\Delta^2 (1 - \rho^2) \pi_d^2}{2n} - \frac{\sigma_{22} (1 - \rho^2) \pi_d^3}{n} - \frac{\sigma_{22} (1 - \rho^2) \pi_d^4}{n}. \quad (3.31)$$

The difference is again notably very small. It matches the difference seen under J2R, plus the additional component, $2\pi_d(\sigma_{11} - 2\sigma_{12})\pi_d/n$. This term is $O(\pi_d^2 n^{-1})$. Since $\pi_d < 1$ this extra term will also always be relatively small in comparison to the information anchoring variance we are targeting, which is $O(n^{-1})$. Thus we see Rubin's variance estimator also provides an excellent approximation of the ideal information anchored variance that preserves the loss of information in the primary analysis under CIR.

The ratio of the difference between Rubin's variance estimator and the information anchored variance (3.31) to the information anchored variance (3.30) under CIR is,

$$\frac{(1 - \rho^2) \pi_d^2 [-\Delta^2 - 2\sigma_{22} (\pi_d + \pi_d^2)] + 4\sigma_{12}^2 \sigma_{11}^{-1} \pi_d^2 + 4(\sigma_{11} - 2\sigma_{12}) \pi_d^2}{4\sigma_{22} + \pi_d [\Delta^2 (2(1 - \pi_d) + (1 - \rho^2) \pi_d) + 2\sigma_{22} (1 - \rho^2) (1 + \pi_d + \pi_d^2 + \pi_d^3)]}.$$

3.2.4 Last mean carried forward

Finally we calculate Rubin's MI variance estimator in the LMCF setting.

Rubin's variance estimator under last mean carried forward

The data distribution under LMCF is formed using the mean from the active arm at time 1 for time 1 and time 2 and variance from the active arm. Appendix B.3 provides the full details of the steps taken to form the conditional model this implies for time 2 given time 1. Again assuming the data is normal, variance is known and improper priors for the missing observations themselves, then the posterior of the model parameters is approximately normal, with prior dominated by the observed data. The LMCF imputation model for patient i and imputation k is,

$$\tilde{Y}_{ai2,k} = \bar{Y}_{a1} + u_k + (r/q + b_k) (Y_{ai1} - \bar{Y}_{a1} - u_k) + e_{i,k}, \text{ for } i \in \mathcal{D}, k = 1, \dots, K, \quad (3.32)$$

where,

$$\begin{aligned} u_k &\sim N(0, n^{-1} \sigma_{11}) \\ b_k &\sim N(0, q^{-1} \sigma_{2.1}) \\ e_{i,k} &\sim N(0, \sigma_{2.1}), \end{aligned}$$

and under LMCF $r = \sum_{i \in \mathcal{O}} (Y_{ai1} - \bar{Y}_{a1,o}) (Y_{ai2} - \bar{Y}_{a2,o})$ and $q = \sum_{i \in \mathcal{O}} (Y_{ai1} - \bar{Y}_{a1,o})^2$. Using the

above imputation model we obtain the treatment effect for the k^{th} imputation over the missing cases as,

$$\hat{\theta}_{\text{LMCF}, k} = \hat{\mu}_{a2,k} - \hat{\mu}_{r2} = \frac{n_o}{n} \bar{Y}_{a2o} + \frac{n_d}{n} \bar{Y}_{a2d,k} - \bar{Y}_{r2},$$

where,

$$\bar{Y}_{a2d,k} = \frac{1}{n_d} \sum_{i \in \mathcal{D}} \tilde{Y}_{ai2,k} = \bar{Y}_{a1} + u_k + \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{a1} - u_k) + \bar{e}_k,$$

and $\bar{e}_k = \frac{1}{n_d} \sum_{i \in \mathcal{D}} e_{i,k}$ and $\bar{Y}_{a1d} = \frac{1}{n_d} \sum_{i \in \mathcal{D}} Y_{ai1}$. Averaging the sample mean estimate for the k^{th} imputation, over the imputation set, K , gives Rubin's MI estimator as,

$$\hat{\theta}_{\text{LMCF, MI}} = \frac{n_o}{n} \bar{Y}_{a2o} + \frac{n_d}{n} \left[\bar{Y}_{a1} + \bar{u} + \left(\frac{r}{q} + \bar{b} \right) (\bar{Y}_{a1d} - \bar{Y}_{a1} - \bar{u}) + \bar{e} \right] - \bar{Y}_{r2},$$

where $\bar{u} = \frac{1}{K} \sum_{k=1}^K u_k$, $\bar{b} = \frac{1}{K} \sum_{k=1}^K b_k$ and $\bar{e} = \frac{1}{K} \sum_{k=1}^K \bar{e}_k$. Taking the necessary expectations reveals,

$$E(\hat{\theta}_{\text{LMCF, MI}}) = \frac{n_o}{n} \mu_{a2} + \frac{n_d}{n} \mu_{a1} - \mu_{r2}.$$

This is unbiased for the treatment effect under LMCF. For Rubin's MI variance estimator under LMCF we require,

$$\begin{aligned} E(\hat{W}) = E(\hat{\sigma}_k^2) = & E\left(\frac{\frac{1}{n-1} \sum_{i=1}^n (Y_{ri2} - \hat{\mu}_{r2})^2}{n} \right) \\ & + E\left(\frac{\frac{1}{n-1} \left[\sum_{i \in \mathcal{O}} (Y_{ai2} - \hat{\mu}_{a2,k})^2 + \sum_{i \in \mathcal{D}} (\hat{Y}_{ai2,k} - \hat{\mu}_{a2,k})^2 \right]}{n} \right). \end{aligned} \quad (3.33)$$

We again define the first component on the RHS of (3.33) as $W1$ and the second component on the RHS of (3.33) as $W2$. For time 1 a baseline where, $\mu_{a1} = \mu_{r1}$, $\pi_d = n_d/n$ and we redefine $\Delta = (\mu_{a2} - \mu_{a1})$ we obtain,

$$E(W1) = \frac{\sigma_{22}}{n},$$

$$E(W2) = \frac{\sigma_{22}}{n} + \frac{\Delta^2 \pi_d n_o}{n(n-1)} + \frac{2\pi_d(1-\pi_d)}{n(n-1)} \left[\sigma_{11} - 2\sigma_{12} + \frac{\sigma_{12}^2}{\sigma_{11}} \right] + \frac{2\sigma_{22}(1-\rho^2)\pi_d}{n(n_o-1)}. \quad (3.34)$$

We now consider $(1 + \frac{1}{K}) E(\hat{B})$. Taking the necessary expectations reveals,

$$\left(1 + \frac{1}{K}\right) E(\hat{B}) = \left(1 + \frac{1}{K}\right) \frac{\pi_d^2}{n} \left[\sigma_{11} - 2\sigma_{12} + \frac{\sigma_{12}^2}{\sigma_{11}} + \sigma_{22}(1-\rho^2) \left(\frac{n_o n + n_d}{n_d(n_o-1)} \right) \right]. \quad (3.35)$$

Assuming that n is sufficiently large so that we may take $(n-1)$ to be n and (n_o-1) to be n_o , gives Rubin's variance estimator asymptotically as,

$$E[V_{\text{LMCF, MI}}] = \frac{2\sigma_{22}}{n} + \frac{\Delta^2 \pi_d(1-\pi_d)}{n} + \frac{2\pi_d(1-\pi_d)}{n^2} \left[\sigma_{11} - 2\sigma_{12} + \frac{\sigma_{12}^2}{\sigma_{11}} \right] + \frac{2\sigma_{22}(1-\rho^2)\pi_d}{n^2(1-\pi_d)} + \left(1 + \frac{1}{K}\right) \frac{\pi_d^2}{n} \left[\sigma_{11} - 2\sigma_{12} + \frac{\sigma_{12}^2}{\sigma_{11}} + \sigma_{22}(1-\rho^2) \left(\frac{1}{\pi_d} + \frac{1}{n(1-\pi_d)} \right) \right].$$

For infinite K this is,

$$E[V_{\text{LMCF, MI}}] = \frac{2\sigma_{22}}{n} + \frac{\Delta^2 \pi_d(1-\pi_d)}{n} + \frac{\sigma_{22}(1-\rho^2)\pi_d}{n} \left[1 + \frac{2}{n(1-\pi_d)} + \frac{\pi_d}{n(1-\pi_d)} \right] + \frac{\pi_d(2+(n-2)\pi_d)}{n^2} \left[\sigma_{11} - 2\sigma_{12} + \frac{\sigma_{12}^2}{\sigma_{11}} \right]. \quad (3.36)$$

Table B.9 summarises Rubin's derived variance under LMCF alongside simulation results. In all settings the derivation results are within 2 MCSE's of the simulation results.

Information anchoring under last mean carried forward

The expectation of the ideal information anchored variance estimate that preserves the fraction of missing information in the primary design based analysis under LMCF, where $\Delta = (\mu_{a2} - \mu_{a1})$ up to the order π_d^4/n terms is,

$$\begin{aligned}
E[V_{\text{anchored}}] &= \frac{2\sigma_{22}}{n} + \frac{\Delta^2 \pi_d (1 - \pi_d)}{n} + \frac{\sigma_{22} (1 - \rho^2) \pi_d}{n} + \left[\frac{\sigma_{22}}{n} + \frac{\Delta^2}{2n} \right] (1 - \rho^2) \pi_d^2 \\
&\quad + \frac{\sigma_{22} (1 - \rho^2) \pi_d^3}{n} + \frac{\sigma_{22} (1 - \rho^2) \pi_d^4}{n}.
\end{aligned} \tag{3.37}$$

We now compare this to Rubin's LMCF MI variance estimator where $K \rightarrow \infty$ (3.36). The difference (3.36) - (3.37) is,

$$\begin{aligned}
E[V_{\text{LMCF, MI}}] - E[V_{\text{anchored}}] &= + \frac{\pi_d (2 + (n - 2)\pi_d)}{n^2} \left[\sigma_{11} - 2\sigma_{12} + \frac{\sigma_{12}^2}{\sigma_{11}} \right] \\
&\quad + \frac{\sigma_{22} (1 - \rho^2) \pi_d (2 + \pi_d)}{n^2 (1 - \pi_d)} - \frac{\Delta^2 (1 - \rho^2) \pi_d^2}{2n} \\
&\quad - \frac{\sigma_{22} (1 - \rho^2) \pi_d^2}{n} - \frac{\sigma_{22} (1 - \rho^2) \pi_d^3}{n} - \frac{\sigma_{22} (1 - \rho^2) \pi_d^4}{n}.
\end{aligned}$$

The difference between Rubin's MI variance and the ideal variance under LMCF is dominated by the following terms which have the largest order,

$$\begin{aligned}
\frac{\pi_d^2}{n} \left[\sigma_{11} - 2\sigma_{12} + \frac{\sigma_{12}^2}{\sigma_{11}} \right] - \frac{\Delta^2 (1 - \rho^2) \pi_d^2}{2n} - \frac{\sigma_{22} (1 - \rho^2) \pi_d^2}{n} - \frac{\sigma_{22} (1 - \rho^2) \pi_d^3}{n} - \frac{\sigma_{22} (1 - \rho^2) \pi_d^4}{n}.
\end{aligned} \tag{3.38}$$

Under LMCF the difference once again depends on the trial size, the amount of missingness in the active arm (n_d), the correlation between the time 1 and time 2 measurements (ρ^2), and the variance of the data at time 2 (σ_{22}), but now additionally depends further on the mean difference between time 1 and time 2 in the active arm ($\Delta = \mu_{a2} - \mu_{a1}$). Rubin's variance estimator will better approximate the information anchored variance with smaller amounts of missingness and when the mean at time 1 is closer to mean 2. However the terms in the difference are of a very small order and in practice negligible. The largest term in (3.38) is $O(\pi_d^2 n^{-1})$. This is tiny in comparison to the information anchored variance, which is $O(n^{-1})$. We see that Rubin's variance estimator also approximates the ideal information anchored variance under LMCF. For larger n the difference approaches zero. Under LMCF the ratio of the difference between Rubin's variance estimator and the information anchored variance (3.38) to the information anchored variance (3.37) is,

$$\frac{(1 - \rho^2) \pi_d^2 [-\Delta^2 - 2\sigma_{22} (1 + \pi_d + \pi_d^2)] + 2\pi_d^2 (\sigma_{11} - 2\sigma_{12} + \sigma_{12}^2 \sigma_{11}^{-1})}{4\sigma_{22} + \pi_d [\Delta^2 (2(1 - \pi_d) + (1 - \rho^2) \pi_d) + 2\sigma_{22} (1 - \rho^2) (1 + \pi_d + \pi_d^2 + \pi_d^3)]}.$$

Our exposition has revealed that in the described baseline and single follow-up setting,

Rubin's variance estimator approximates the information anchored variance that preserves the loss of information in the design based primary analysis for all reference based MI sensitivity analysis scenarios.

We summarise the results of the derivations undertaken in this Section in Tables 3.2 and 3.3. The difference between Rubin's variance and the information anchored variance is of a small order in each case. Rubin's variance estimator keeps the information constant across the primary and sensitivity analysis as desired. We now explore this property in a simulation study based on real data.

Imputation Method	Rubin's MI variance - Information anchored variance (largest order terms)
CR	$-\frac{\Delta^2(1-\rho^2)\pi_d^2}{2n} - \frac{\sigma_{22}(1-\rho^2)\pi_d^3}{n} - \frac{\sigma_{22}(1-\rho^2)\pi_d^4}{n}$
J2R	$-\frac{\Delta^2(1-\rho^2)\pi_d^2}{2n} - \frac{\sigma_{22}(1-\rho^2)\pi_d^3}{n} - \frac{\sigma_{22}(1-\rho^2)\pi_d^4}{n} + 2\frac{\sigma_{12}^2}{\sigma_{11}}\frac{\pi_d^2}{n}$
CIR	$-\frac{\Delta^2(1-\rho^2)\pi_d^2}{2n} - \frac{\sigma_{22}(1-\rho^2)\pi_d^3}{n} - \frac{\sigma_{22}(1-\rho^2)\pi_d^4}{n} + 2\frac{\sigma_{12}^2}{\sigma_{11}}\frac{\pi_d^2}{n} + 2(\sigma_{11} - 2\sigma_{12})\frac{\pi_d^2}{n}$
LMCF	$-\frac{\Delta^2(1-\rho^2)\pi_d^2}{2n} - \frac{\sigma_{22}(1-\rho^2)\pi_d^2}{n} - \frac{\sigma_{22}(1-\rho^2)\pi_d^3}{n} - \frac{\sigma_{22}(1-\rho^2)\pi_d^4}{n} + \frac{\pi_d^2}{n} \left[\sigma_{11} - 2\sigma_{12} + \frac{\sigma_{12}^2}{\sigma_{11}} \right]$

Table 3.2: Difference between Rubin's MI variance estimator and the information anchored variance for CR, J2R, CIR and LMCF imputation with baseline and a single follow-up. Note: $\Delta = \mu_{a2} - \mu_{r2}$ for CR, J2R and CIR. For LMCF $\Delta = \mu_{a2} - \mu_{a1}$.

Imputation Method	(Rubin's - Information anchored variance)/Information anchored variance (largest order terms)
CR	$\frac{(1-\rho^2)\pi_d^2[-\Delta^2 - 2\sigma_{22}(\pi_d + \pi_d^2)]}{4\sigma_{22} + \pi_d[\Delta^2(2(1-\pi_d) + (1-\rho^2)\pi_d) + 2\sigma_{22}(1-\rho^2)(1 + \pi_d + \pi_d^2 + \pi_d^3)]}$
J2R	$\frac{(1-\rho^2)\pi_d^2[-\Delta^2 - 2\sigma_{22}(\pi_d + \pi_d^2)] + 4\sigma_{12}^2\sigma_{11}^{-1}\pi_d^2}{4\sigma_{22} + \pi_d[\Delta^2(2(1-\pi_d) + (1-\rho^2)\pi_d) + 2\sigma_{22}(1-\rho^2)(1 + \pi_d + \pi_d^2 + \pi_d^3)]}$
CIR	$\frac{(1-\rho^2)\pi_d^2[-\Delta^2 - 2\sigma_{22}(\pi_d + \pi_d^2)] + 4\sigma_{12}^2\sigma_{11}^{-1}\pi_d^2 + 4(\sigma_{11} - 2\sigma_{12})\pi_d^2}{4\sigma_{22} + \pi_d[\Delta^2(2(1-\pi_d) + (1-\rho^2)\pi_d) + 2\sigma_{22}(1-\rho^2)(1 + \pi_d + \pi_d^2 + \pi_d^3)]}$
LMCF	$\frac{(1-\rho^2)\pi_d^2[-\Delta^2 - 2\sigma_{22}(1 + \pi_d + \pi_d^2)] + 2\pi_d^2(\sigma_{11} - 2\sigma_{12} + \sigma_{12}^2\sigma_{11}^{-1})}{4\sigma_{22} + \pi_d[\Delta^2(2(1-\pi_d) + (1-\rho^2)\pi_d) + 2\sigma_{22}(1-\rho^2)(1 + \pi_d + \pi_d^2 + \pi_d^3)]}$

Table 3.3: The ratio of the difference between Rubin's MI variance estimator and the information anchored variance to the information anchored variance for CR, J2R, CIR and LMCF imputation with baseline and a single follow-up. Note: $\Delta = \mu_{a2} - \mu_{r2}$ for CR, J2R and CIR. For LMCF $\Delta = \mu_{a2} - \mu_{a1}$.

3.3 Simulation study to investigate Rubin’s variance estimator

We present the results of a simulation study conducted to further evaluate the information anchoring properties of Rubin’s MI variance estimator in a variety of trial settings for each of the reference based procedures (CR, J2R, CIR and LMCF). The baseline and week 12 data from the asthma trial described in Subsection 1.7.2 formed the motivating example for generating the datasets.

3.3.1 Methods

Baseline (time 1) and follow-up (time 2) data were generated using an underlying bivariate normal distribution with means and covariance matrices similar to those obtained in the asthma study at baseline and week 12 as,

$$\mu_{\text{placebo}} = [2, 1.9], \mu_{\text{active}} = [2, \mu_{a,2}],$$

$$\Sigma_{\text{placebo}} = \Sigma_{\text{active}} = \begin{bmatrix} 0.4 & 0.2 \\ 0.2 & 0.6 \end{bmatrix}.$$

In the asthma study a mean treatment group difference of $\Delta = 0.3$ was observed at follow-up ($\mu_{a,2} = 2.2$). We consider a range of $\mu_{a,2}$ corresponding to treatment differences of $\Delta = \{0, 0.3, 1\}$ for 10–70% missing data in the active arm only at time 2.

One thousand independent complete datasets were generated for each combination of $\mu_{a,2}$ and missingness, and a sample size of $n=500$ patients ($n=250$ placebo, $n=250$ active) was drawn for all simulations.

Interest lies in the mean treatment group difference at time 2. For each setting, the treatment effect and estimated variance were computed using the full data. The analysis model was a linear regression of the time 2 response on randomised treatment group. Within the active arm, a MCAR missingness mechanism, was then imposed on the FEV₁ outcome at time 2 by independently sampling a binary indicator from a Bernoulli distribution with success probability equal to the proportion of missing data required for patients in the active arm. If the indicator equalled one then the corresponding outcome at time 2 was set missing.

The reference based MI approach of Carpenter, Roger and Kenward [13] was then applied to each incomplete dataset under MAR, CR, J2R, CIR and LMCF with 50 imputations. We used the same set of 1000 incomplete datasets for each imputation procedure in order to minimise the Monte-Carlo variability in our comparisons.

In each scenario post-deviation data were regenerated using the appropriate conditional normal distributions in order to compute the treatment effect and its variance had the post-deviation data been observed in the particular de-facto setting ($V_{\text{DF, full}}$).

The main outcomes of interest in each scenario were Rubin’s MI estimate of variance and the information anchored variance, computed using (3.2), for the treatment effect at time 2. Estimates were averaged over the 1000 simulations in each scenario for comparison.

Additional simulations were run with the above mean and variance parameters but a smaller ($n=100$) and larger ($n=1000$) sample size per arm to explore the effect the trial size has on the information anchoring properties of Rubin’s variance estimator. We then considered the effect of having a smaller or larger conditional variance at time 2 given baseline for the original mean structure and sample size of $n=250$ per arm, with the following alternative covariance structures for each missingness scenario,

Low covariance set-up,

$$\Sigma_{\text{placebo,low}} = \Sigma_{\text{active,low}} = \begin{bmatrix} 1 & 0.05 \\ 0.05 & 1 \end{bmatrix}.$$

High covariance variance set-up,

$$\Sigma_{\text{placebo,high}} = \Sigma_{\text{active,high}} = \begin{bmatrix} 1 & 0.75 \\ 0.75 & 1 \end{bmatrix}.$$

3.3.2 Results

Simulations with a moderate treatment effect of $\Delta = 0.3$ ($\mu_{a,2} = 2.2$), as observed in the asthma trial, show excellent information anchoring by Rubin’s variance estimator for up to 40–50% missing data in the active arm for all reference based settings (Figure 3.1) i.e. for realistic proportions of missing data.

Rubin’s variance estimator reflects our increased uncertainty with greater proportions of missing data and is appropriately greater than the variance estimator based on completely observed data under each postulated scenario.

This is in contrast to the behaviour of the empirical long-run sampling variance of the reference based MI estimator (the variance of the MI treatment estimator over the 1000 simulations), which is equivalent to the long-run sampling variance of the reference based ML estimator, labelled Sampling MI/ML variance in Figures 3.1 to 3.5. In all reference based settings the sampling variance of the MI/ML estimator does not increase as the proportion of missing data increases to reflect the loss of information. Specifically for CR, CIR and J2R we see the sampling variance of the MI/ML estimate tends towards zero as the proportion of missing data increases. This is because the mean proposed for the active deviators in imputation is no longer independent of the mean from the reference group. As established in Subsection 2.2.2, when looking at the difference between these two groups the introduced covariance is then taken away in the sampling variance. The sampling variance of the reference based MI/ML estimator is smaller than the variance estimator based on completely observed reference based deviation data, thus is not appropriate.

For smaller and larger treatment effects we see the same behaviour (see Figures C.1 to and C.2). Even with a large treatment effect of $\Delta = 1.0$ ($\mu_{a,2} = 2.9$) corresponding to $\frac{5}{3}\sigma_{22}$ —thus much

larger than would be anticipated in any trial in practice— there is excellent agreement between Rubin’s variance estimator and the information anchored variance for up to 40–50% missing data in the active arm.

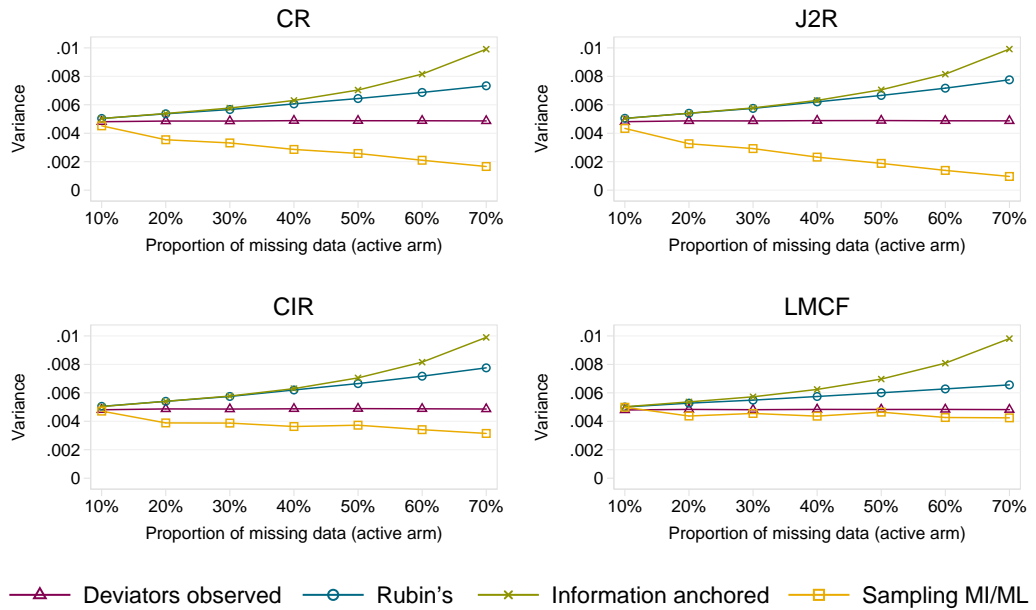


Figure 3.1: Rubin’s variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with the asthma trial treatment effect of $\Delta = 0.3$ and $n=250$ per arm.

When a smaller sample size of $n=100$ per arm is considered we still see an excellent approximation between Rubin’s variance estimator and the information anchored variance for up to 40–50% missing data, see Figure 3.2 with $\Delta = 0.3$ (for $\Delta = 0$ and $\Delta = 1$ with $n=100$ per arm see Figures C.3 and C.4). When the sample size was inflated, see Figure 3.3 with $\Delta = 0.3$, as expected we also see excellent agreement and smaller differences (for $\Delta = 0$ and $\Delta = 1$ with $n=1000$ per arm see Figures C.5 and C.6).

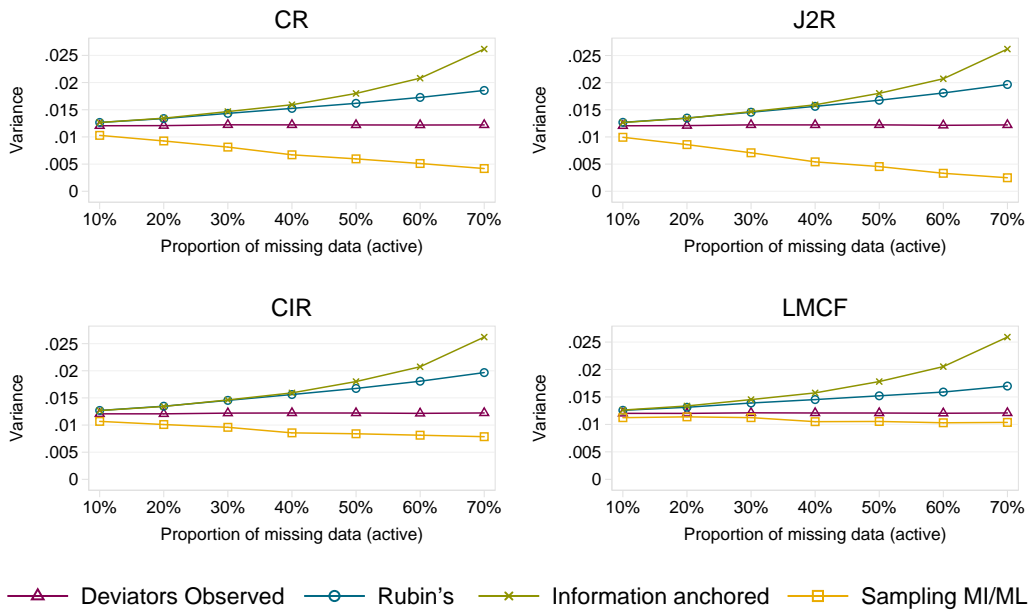


Figure 3.2: Rubin's variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with the asthma trial treatment effect of $\Delta = 0.3$ and $n=100$ per arm.

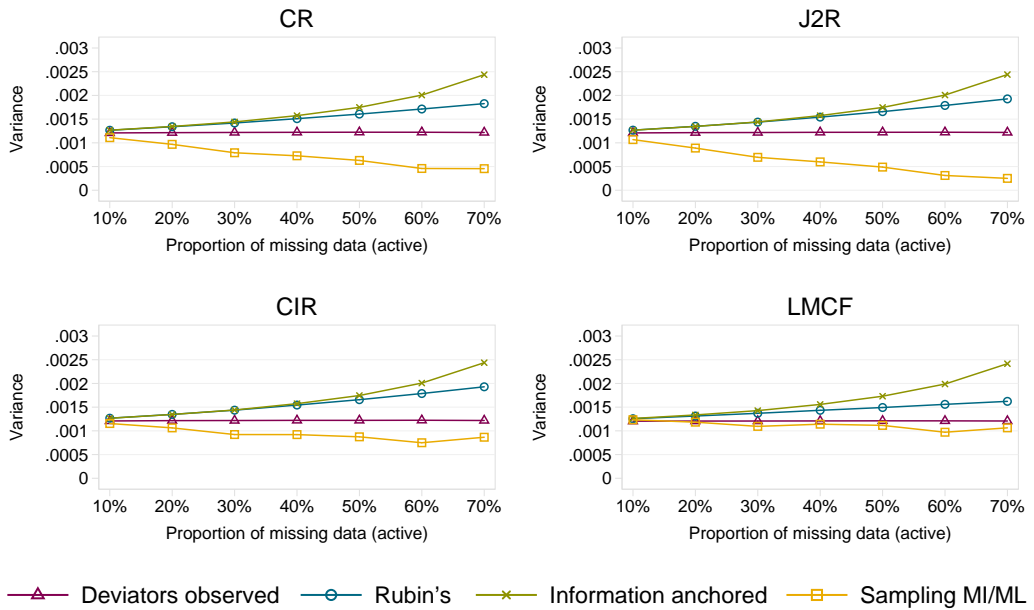


Figure 3.3: Rubin's variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with the asthma trial treatment effect of $\Delta = 0.3$ and $n=1000$ per arm.

We now consider more extreme covariance structures. We see a similar relationship between Rubin's variance estimator and the ideal information anchored variance when we have a low covariance between time 1 and 2, i.e. a larger conditional variance (see Figure 3.4 with $\Delta = 0.3$; for $\Delta = 0$ and $\Delta = 1$ see Figures C.7 and C.8). There is still excellent agreement for up to 40–50% missingness in the active arm despite a smaller correlation between time 1 and time 2 outcomes. We see when we have a higher covariance between time 1 and 2, i.e. a smaller conditional variance agreement is

even better and excellent for higher (60–70%) missingness (see Figure 3.5 with $\Delta = 0.3$; for $\Delta = 0$ and $\Delta = 1$ see Figures C.9 and C.10).

The simulation results support the results from the derivations. That is, Rubin’s variance estimator provides an excellent approximation for the information anchored variance that preserves the information loss in the primary design based analysis.

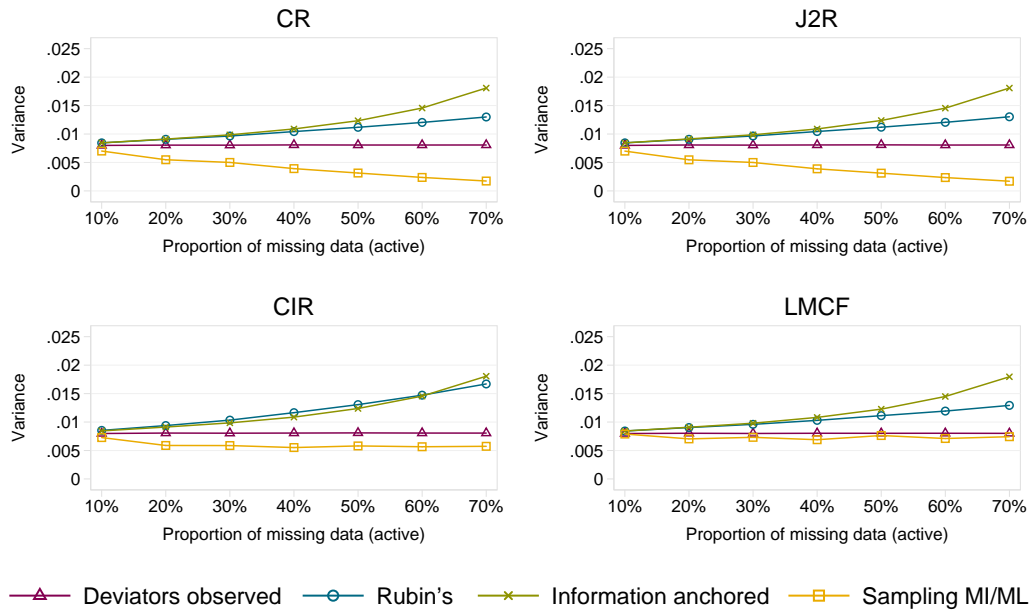


Figure 3.4: Rubin’s variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with the asthma trial treatment effect of $\Delta = 0.3$ and a low covariance between time 1 and time 2 outcomes ($\sigma_{12} = 0.05$). $n=250$ per arm.

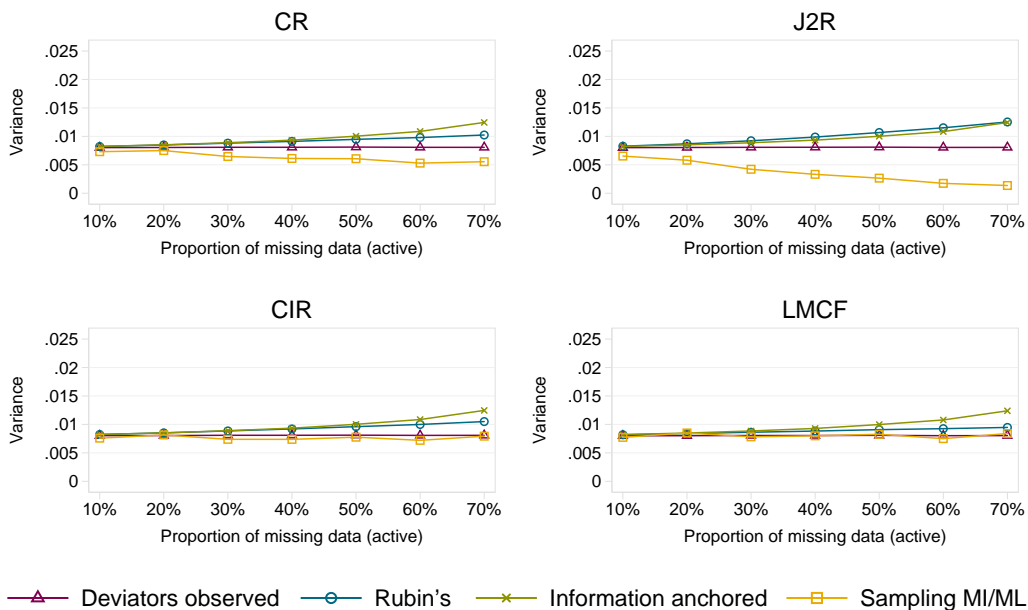


Figure 3.5: Rubin’s variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with the asthma trial treatment effect of $\Delta = 0.3$ and a high covariance between time 1 and time 2 outcomes ($\sigma_{12} = 0.75$). $n=250$ per arm.

3.3.3 Discussion

So far we have seen that $\hat{\theta}_{DF, MI}$ is an unbiased estimator, whose long-run sampling variance is too small. But whose variance can be estimated under the information anchoring principle using Rubin’s MI rules. We have shown that Rubin’s variance estimator approximately preserves the loss of information seen in the primary design based analysis across reference based sensitivity analysis in a variety of realistic trial settings. When considering more extreme scenarios, i.e. altering the treatment effect and covariance structure we see Rubin’s variance estimator still provides an excellent approximation for the information anchoring variance for proportions of missing data realistically seen. This provides a solid justification for using Rubin’s variance formula to provide an estimate of variance in the reference based sensitivity analysis settings.

The approximation between Rubin’s variance and the required information anchored variance is notably not so good when missingness is greater than 50% in the active arm. With 50% or more missing data we would indeed be cautious of results from any method of analysis. We would be worried why our trial arm has so much missing data and would need to carefully investigate the failures of the trial further.

3.4 Baseline adjusted setting

We now investigate Rubin’s MI variance estimator for the baseline adjusted treatment effect. We consider the same baseline and single follow-up trial set-up outlined at the beginning of this chapter. Interest now lies in the baseline adjusted mean difference between the treatment groups

at the follow-up time point, with an appropriate measure of precision.

3.4.1 Information anchoring for the baseline adjusted treatment estimator

As in the unadjusted setting, in reference based sensitivity analysis it is postulated that the appropriate variance estimate for the baseline adjusted treatment effect is the design based variance estimator we would obtain if we observed the deviation data in the given de-facto scenario (one of CR, J2R, CIR or LMCF), inflated by the loss of information seen in the primary design based analysis (under MAR). We want to keep the information loss constant across the primary and sensitivity analysis. As discussed in Section 3.1 this is a natural and sensible principled approach. That is, the information anchored variance estimator required for the treatment estimate in sensitivity analysis is,

$$V_{\text{anchored}} = \frac{V_{\text{DJ, MI}}}{V_{\text{DJ, full}}} \times V_{\text{DF, full}}.$$

In the adjusted treatment effect setting, $V_{\text{DJ, full}}$ denotes the variance estimator for the baseline adjusted treatment estimator, from the design based analysis with full data (no deviation); $V_{\text{DJ, MI}}$ denotes Rubin's variance estimator for the baseline adjusted treatment estimator from the design based MAR MI analysis (primary analysis); and $V_{\text{DF, full}}$ denotes the design based variance estimator for the baseline adjusted treatment estimator if post-deviation data were observed in the postulated reference based scenario (one of CR, J2R, CIR or LMCF). We now compute the expectation of these components in the baseline and single follow-up setting, in order to compare the desired information anchored variance with the expectation of Rubin's variance estimator. It is trivial to see from the unadjusted calculation in Section 3.2 that all MI estimates are unbiased under their scenario, so we don't discuss this further.

To identify the expected value of the variance estimate for the baseline adjusted treatment estimate in full data, when no deviations occur, $V_{\text{DJ, full}}$, we consider the conditional distributions of $\mathbf{Y}_{z2}|\mathbf{Y}_{z1}^*$,

$$(\mathbf{Y}_{z2}|\mathbf{Y}_{z1}^*) \sim N(\mathbf{X}_z\beta_z; \sigma_{2.1,z}),$$

where \mathbf{Y}_{z2} is the vector of outcome data at time 2 from the n patients in arm z , $\mathbf{Y}_{z2} = (Y_{z12}, \dots, Y_{zn2})^T$ for $z = a, r$. To simplify calculations \mathbf{Y}_{z1}^* is the vector of baseline data from the n patients in arm z after adjustment for the overall mean at baseline in arm z (so \mathbf{Y}_{z1}^* has mean zero). That is, $\mathbf{Y}_{z1}^* = (Y_{z11}^*, \dots, Y_{zn1}^*)^T$, where $Y_{zi1}^* = Y_{zi1} - \bar{Y}_{z1}$ and $\bar{Y}_{z1} = \frac{1}{n} \sum_{i=1}^n Y_{zi1}$. \mathbf{X}_z is the design matrix,

$$\mathbf{X}_z = \begin{bmatrix} 1 & Y_{z11}^* \\ 1 & Y_{z21}^* \\ \vdots & \vdots \\ 1 & Y_{zn1}^* \end{bmatrix}.$$

The usual least square estimates are given by $(\mathbf{X}_z^T \mathbf{X}_z)^{-1} \mathbf{X}_z^T \mathbf{Y}_{z2}$ which gives,

$$\hat{\beta}_z = \left(\frac{\bar{Y}_{z2}}{\frac{\sum_{i=1}^n Y_{zi1}^* Y_{zi2}}{\sum_{i=1}^n Y_{zi1}^{*2}}} \right).$$

Having adjusted Y_{z1} for the overall mean at baseline in that arm (\bar{Y}_{z1}), \bar{Y}_{z2} represents the average level of Y_{z2} for a patient with average Y_{z1} . The variance of the least square estimates are,

$$(\mathbf{X}_z^T \mathbf{X}_z)^{-1} \sigma_{2.1,z} = \begin{bmatrix} n^{-1} & 0 \\ 0 & \frac{1}{\sum_{i=1}^n Y_{zi1}^{*2}} \end{bmatrix} \sigma_{2.1,z},$$

where $\sigma_{2.1,z}$ is the residual variance in arm z , which can be calculated as, $\sigma_{2.1,z} = \sigma_{22,z} - \sigma_{12,z}^2 / \sigma_{11,z}$.

In the current set-up, using the definitions above, the adjusted treatment group difference of interest, denoted $\hat{\theta}_{\text{DJ, full}}$, is calculated as, $\bar{Y}_{a2} - \bar{Y}_{r2}$. The expectation of this is, $\mu_{a2} - \mu_{r2}$. This is the effect for an active patient with average baseline - reference patient with average baseline. The variance estimator for this adjusted treatment difference is, $\hat{\sigma}_{2.1,a}/n + \hat{\sigma}_{2.1,r}/n$.

When no deviations occur, conditional on time 1, the expected value of the residual variance terms are, $E[\hat{\sigma}_{2.1,a}] = E[\hat{\sigma}_{2.1,r}] = \sigma_{2.1} = \sigma_{22} - \sigma_{12}^2 / \sigma_{11}$. The expected value of the variance estimate for the adjusted treatment effect when no deviations are observed is therefore,

$$E[V_{\text{DJ, full}}] = \frac{2\sigma_{2.1}}{n}. \quad (3.39)$$

To identify the expected value of the variance estimate for the baseline adjusted treatment estimate when deviations occur in the active arm and the post-deviation data are observed, in the de-facto reference based settings, $V_{\text{DF, full}}$, we reconsider the conditional distribution of $\mathbf{Y}_{a2} | \mathbf{Y}_{a1}^*$ in the presence of deviation. \mathbf{Y}_{a2} is now the vector of outcome data at time 2 from the active arm including deviators and non-deviators. \mathbf{Y}_{a1}^* is the vector of baseline data from the active arm. After adjustment for the overall mean at baseline in the active arm.

$$(\mathbf{Y}_{a2} | \mathbf{Y}_{a1}^*) \sim N(\mathbf{X}_a \beta_a; \sigma_{2.1,a}).$$

Here \mathbf{X}_a is the design matrix,

$$\mathbf{X}_a = \begin{bmatrix} 1 & Y_{a11}^* \\ 1 & Y_{a21}^* \\ \vdots & \vdots \\ 1 & Y_{an1}^* \end{bmatrix}.$$

The usual least square estimates are given by $(\mathbf{X}_a^T \mathbf{X}_a)^{-1} \mathbf{X}_a^T \mathbf{Y}_{a2}$ which gives,

$$\hat{\beta}_a = \left(\frac{\frac{1}{n}(n_o \bar{Y}_{a2o} + n_d \bar{Y}_{a2d})}{\frac{\sum_{i=1}^n Y_{ai1}^* Y_{ai2}}{\sum_{i=1}^n Y_{ai1}^{*2}}} \right),$$

with variance,

$$(X_a^T X_a)^{-1} \sigma_{2.1,a} = \left| \begin{array}{cc} n^{-1} & 0 \\ 0 & \frac{1}{\sum_{i=1}^n Y_{ai1}^{*2}} \end{array} \right| \sigma_{2.1,a},$$

where $\sigma_{2.1,a}$ is the residual variance, which can be calculated as, $\sigma_{2.1,a} = \sigma_{22,a} - \sigma_{12,a}^2 / \sigma_{11,a}$.

With deviation in the active arm, the adjusted treatment group difference of interest, $\hat{\theta}_{\text{DF, full}}$, is calculated as $(1/n)(n_o \bar{Y}_{a2o} + n_d \bar{Y}_{a2d}) - \bar{Y}_r$. The expectation of this is $(n_o/n)\mu_{a2} + (n_d/n)\mu_{d2} - \mu_{r2}$. The expected value of the variance at time 2 as derived in Appendix B.1 is, $E[\hat{\sigma}_{22,a}] = \sigma_{22} + n_d n_o \Delta^2 / n^3$ and as the design based analysis conditions on baseline data $E[\hat{\sigma}_{12,a}^2 / \hat{\sigma}_{11,a}] = \sigma_{12}^2 / \sigma_{11}$ (assuming $\text{COV}(Y_{ai1}, Y_{ai2}) = \sigma_{12}$ for $i \in \mathcal{O}, \mathcal{D}$). The expected value of the residual variance in the active arm, where $\Delta = \mu_{a2} - \mu_{r2}$, is therefore,

$$E[\hat{\sigma}_{2.1,a}] = \sigma_{22} + \frac{n_d n_o \Delta^2}{n^3} - \frac{\sigma_{12}^2}{\sigma_{11}} = \sigma_{2.1} + \frac{n_d n_o \Delta^2}{n^3}.$$

The expected value of the variance estimate for the adjusted treatment effect when reference based post-deviation data is observed is therefore,

$$E[V_{\text{DF, full}}] = E\left[\frac{\hat{\sigma}_{2.1,r}}{n}\right] + E\left[\frac{\hat{\sigma}_{2.1,a}}{n}\right] = \frac{2\sigma_{2.1}}{n} + \frac{\Delta^2 n_d n_o}{n^3}. \quad (3.40)$$

The last component needed to establish the expected value of the required variance estimate for the baseline adjusted treatment effect is the expected value of Rubin's variance estimator for the

baseline adjusted treatment effect in the primary MI MAR setting.

We follow the same set-up and initial developments for an appropriate imputation distribution under MAR as in the unadjusted treatment effect setting outlined in Subsection 3.1.1. The imputation model is as in (3.3).

The MI estimator, calculated using Rubin's rules is, $\hat{\theta}_{\text{DJ, MI}} = \frac{1}{K} \sum_{k=1}^K \hat{\theta}_{\text{DJ, } k}$, where $\hat{\theta}_{\text{DJ, } k}$ represents the baseline adjusted treatment estimator in the k^{th} imputed dataset. From the full data setting outlined above we infer that the baseline adjusted treatment estimator, in imputation k is computed as, $\hat{\theta}_{\text{DJ, } k} = \bar{Y}_{a2,k} - \bar{Y}_{r2}$. The expectation of $\bar{Y}_{a2,k}$ and \bar{Y}_{r2} have already been computed under MAR in Subsection 3.1.1, which reveals $E[\hat{\theta}_{\text{DJ, } k}] = \mu_{a2} - \mu_{r2}$ and $E[\hat{\theta}_{\text{DJ, MI}}] = \mu_{a2} - \mu_{r2}$. This is unbiased under the described conditions, where adjustment for the overall mean at baseline in each arm has been made, prior to including this variable in the analysis.

Our target is Rubin's variance estimator for the baseline adjusted treatment effect, which is given by, $\hat{W} + (1 + \frac{1}{K})\hat{B}$, where $\hat{W} = \frac{1}{K} \sum_{k=1}^K \hat{\sigma}_k^2$ and $\hat{\sigma}_k^2$ now represents the variance estimator for the adjusted treatment estimator in imputation k .

We infer from the full data setting outlined above, that the variance estimator for the adjusted treatment difference for imputation k will be, $\hat{\sigma}_k^2 = \hat{\sigma}_{2.1,r}/n + \hat{\sigma}_{2.1,a,k}/n$. Thus,

$$E[\hat{W}] = \frac{\sigma_{2.1}}{n} + \frac{E[\hat{\sigma}_{2.1,a,k}]}{n}.$$

Breaking this down for the active arm, since we are conditioning on baseline (time 1) $E[\hat{\sigma}_{12,a,k}/\hat{\sigma}_{11,a,k}] = \sigma_{12}^2/\sigma_{11}$ we have,

$$E[\hat{\sigma}_{2.1,a,k}] = E[\hat{\sigma}_{22,a,k}] - \frac{\sigma_{12}^2}{\sigma_{11}},$$

where $\hat{\sigma}_{22,a,k}$ denotes the variance of the data at time 2 in the active arm, in the k^{th} imputed dataset. The expectation of $\hat{\sigma}_{22,a,k}$ has already been computed under MAR in Section 3.1.1 as,

$$E[\hat{\sigma}_{22,a,k}] = \sigma_{22} + \frac{2n_d\sigma_{2.1}}{(n_o - 1)(n - 1)}.$$

The between imputation variance is given by $\hat{B} = \frac{1}{K-1} \sum_{k=1}^K (\hat{\theta}_{\text{DJ, } k} - \hat{\theta}_{\text{DJ, MI}})^2$. Taking the necessary expectations, as done in Section 3.1.1 reveals,

$$E[\hat{B}] = \frac{\sigma_{2.1}n_d}{n(n_o - 1)}.$$

Thus the overall expected value of the variance estimator of the MI estimator under MAR, where we assume that n is sufficiently large so that we may take $(n - 1)$ to be n and $(n_o - 1)$ to be n_o is,

$$E[V_{\text{DJ, MI}}] = \frac{2\sigma_{2.1}}{n} + \frac{2\sigma_{2.1}\pi_d}{n^2(1 - \pi_d)} + \left(1 + \frac{1}{K}\right) \frac{\sigma_{2.1}\pi_d}{n(1 - \pi_d)}.$$

For infinite K this is equal to,

$$E[V_{\text{DJ, MI}}] = \frac{2\sigma_{2.1}}{n} + \frac{2\sigma_{2.1}\pi_d}{n^2(1 - \pi_d)} + \frac{\sigma_{2.1}\pi_d}{n(1 - \pi_d)} = \sigma_{2.1} \left[\frac{1}{n} + \frac{1}{n_o} \right] + \frac{2\sigma_{2.1}\pi_d}{n^2(1 - \pi_d)}. \quad (3.41)$$

Table B.2 summarises Rubin's derived variance estimator under MAR alongside simulation results. Derived results are within 2 MCSE's of the simulation results. Rubin's MAR variance estimator is asymptotically equivalent to the expected value of the estimated variance of the baseline adjusted treatment effect in the observed data from a full likelihood analysis, $E[V_{\text{DJ, ML}}]$. That is the effect of an active patient with average baseline - reference patient with average baseline, estimated from a regression of the time 2 outcome on mean adjusted baseline in each arm, for individuals observed at both time points only. When deviation data is missing at time 2 only, under MAR (or MCAR), the active cases observed a time 1 only contribute no information towards the baseline adjusted treatment effect.

To see this we now derive the ML variance. Consider the conditional distribution of $\mathbf{Y}_{a2o}|\mathbf{Y}_{a1o}^*$. \mathbf{Y}_{a2o} is the vector of outcome data at time 2 from the n_o active patients observed at both time points, $\mathbf{Y}_{a2o} = \{Y_{ai2}; i \in \mathcal{O}\}^T$. \mathbf{Y}_{a1o}^* is the vector of baseline data from the non-deviating active patients, after adjustment for the mean baseline outcome at time 1 for the n_o active arm patients, $\mathbf{Y}_{a1o}^* = \{Y_{ai1}^*; i \in \mathcal{O}\}^T$, where $Y_{ai1}^* = Y_{ai1} - \bar{Y}_{a1o}$, $\bar{Y}_{a1o} = \frac{1}{n_o} \sum_{i \in \mathcal{O}} Y_{ai1}$,

$$(\mathbf{Y}_{a2o}|\mathbf{Y}_{a1o}^*) \sim N(\mathbf{X}_{\mathbf{ao}}\beta_{\mathbf{ao}}; \sigma_{2.1,ao}),$$

where $\mathbf{X}_{\mathbf{ao}}$ is the design matrix,

$$\mathbf{X}_{\mathbf{ao}} = \begin{bmatrix} 1 & Y_{a11}^* \\ 1 & Y_{a21}^* \\ \vdots & \vdots \\ 1 & Y_{an_o1}^* \end{bmatrix}.$$

The usual least square estimates are given by $(\mathbf{X}_{\mathbf{ao}}^T \mathbf{X}_{\mathbf{ao}})^{-1} \mathbf{X}_{\mathbf{ao}}^T \mathbf{Y}_{\mathbf{a2o}}$ which gives,

$$\hat{\beta}_{ao} = \left(\begin{array}{c} \bar{Y}_{a2o} \\ \frac{\sum_{i \in \mathcal{O}} Y_{ai1}^* Y_{ai2}}{\sum_{i \in \mathcal{O}} Y_{ai1}^{*2}} \end{array} \right),$$

with variance,

$$(X_{ao}^T X_{ao})^{-1} \sigma_{2.1,ao} = \left| \begin{array}{cc} n_o^{-1} & 0 \\ 0 & \frac{1}{\sum_{i \in \mathcal{O}} Y_{ai1}^{*2}} \end{array} \right| \sigma_{2.1,ao},$$

where $\sigma_{2.1,ao}$ is the residual variance, which can be calculated as, $\sigma_{2.1,ao} = \sigma_{22,ao} - \sigma_{12,ao}^2 / \sigma_{11,ao}$.

The adjusted treatment group difference of interest, under the described conditions, is calculated as, $\bar{Y}_{a2o} - \bar{Y}_r$. The expectation of this is, $\mu_{a2} - \mu_{r2}$. The variance estimator for the adjusted treatment difference will be, $\hat{\sigma}_{2.1,ao}/n_o + \hat{\sigma}_{2.1,r}/n$.

Since conditioning on baseline (time 1) $E[\hat{\sigma}_{2.1,ao}] = E[\hat{\sigma}_{2.1,r}] = \sigma_{2.1} = \sigma_{22} - \sigma_{12}^2 / \sigma_{11}$, the expected value of the variance for the adjusted treatment effect when post-deviation data are unobserved, from a full likelihood analysis is,

$$E[V_{\text{DJ, ML}}] = \sigma_{2.1} \left[\frac{1}{n} + \frac{1}{n_o} \right].$$

This is asymptotically equivalent to the MI MAR variance estimator using Rubin's variance formula (3.41) (for $n \rightarrow \infty, K \rightarrow \infty$). We now have the inflation factor to apply to the design based variance estimator for the adjusted treatment effect where reference based deviation data is observed (3.40) to derive the required variance estimator in sensitivity analysis with missing post-deviation data. That is (3.41) (up to the $O(n^{-2})$ terms) divided by (3.39),

$$\frac{E[V_{\text{DJ, MI}}]}{E[V_{\text{DJ, full}}]} = 1 + \frac{\pi_d}{2(1 - \pi_d)} = 1 + \frac{\pi_d + \pi_d^2 + \pi_d^3 + \pi_d^4 + \dots}{2}.$$

We will now assess whether Rubin's variance formula comes close to the desired information anchored variance estimate in each of the reference based settings.

3.4.2 Rubin's variance estimator for the baseline adjusted treatment estimator

Here we analytically derive Rubin's MI variance estimator in each of the CR, J2R, CIR and LMCF settings and compare this to the information anchored variance estimate. We use the results

from Section 3.2 within computations. In each de-facto scenario, MI proceeds in the same way as when the unadjusted treatment estimator was of interest in Section 3.2. It is just the substantive primary analysis model that is different, because it is now baseline adjusted. Under the described conditions, the adjusted treatment estimate is the effect for an active patient with average baseline - reference patient with average baseline.

Rubin's variance estimator under copy reference

We follow the same set-up and initial developments for an appropriate imputation distribution under CR as in the unadjusted treatment effect setting outlined in Subsection 3.2.1. The imputation model is as in (3.9).

The MI estimator, calculated using Rubin's rules is, $\hat{\theta}_{\text{CR, MI}} = \frac{1}{K} \sum_{k=1}^K \hat{\theta}_{\text{CR, } k}$, where $\hat{\theta}_{\text{CR, } k}$ represents the baseline adjusted treatment estimator in the k^{th} imputed dataset. From the full data setting outlined above we infer that the baseline adjusted treatment estimator in imputation k is computed as, $\hat{\theta}_{\text{CR, } k} = \bar{Y}_{a2,k} - \bar{Y}_{r2}$. The expectation of $\bar{Y}_{a2,k}$ and \bar{Y}_{r2} have already been computed under CR in Subsection 3.2.1, which reveals $E[\hat{\theta}_{\text{CR, } k}] = (n_o/n)(\mu_{a2} - \mu_{r2})$ and $E[\hat{\theta}_{\text{CR, MI}}] = (n_o/n)(\mu_{a2} - \mu_{r2})$.

Rubin's variance estimator for the adjusted treatment effect, is given by, $\hat{W} + (1 + \frac{1}{K})\hat{B}$, where $\hat{W} = \frac{1}{K} \sum_{k=1}^K \hat{\sigma}_k^2$ and $\hat{\sigma}_k^2$ represents the variance estimator for the adjusted treatment difference in imputation k . We infer from the full data setting outlined above, that the variance estimator for the adjusted treatment difference for imputation k will be, $\hat{\sigma}_k^2 = \hat{\sigma}_{2.1,r}/n + \hat{\sigma}_{2.1,a,k}/n$. Thus,

$$E[\hat{W}] = \frac{\sigma_{2.1}}{n} + \frac{E[\hat{\sigma}_{2.1,a,k}]}{n}.$$

Under CR imputation, since we are conditioning on baseline (time 1) $E[\hat{\sigma}_{12,a,k}^2/\hat{\sigma}_{11,a,k}] = \sigma_{12}^2/\sigma_{11}$. We therefore have, $E[\hat{\sigma}_{2.1,a,k}] = E[\hat{\sigma}_{22,a,k}] - \sigma_{12}^2/\sigma_{11}$, where $\hat{\sigma}_{22,a,k}$ denotes the estimated variance of the data at time 2 in the active arm, in the k^{th} imputed dataset. The expectation of $\hat{\sigma}_{22,a,k}$ has already been computed under CR in Subsection 3.2.1 as,

$$E[\hat{\sigma}_{22,a,k}] = \sigma_{22} + \frac{\Delta^2 \pi_d n_o}{(n-1)} + \frac{2\sigma_{2.1} \pi_d}{(n-1)} \left[2 - \pi_d \left(1 + \frac{1}{n} \right) \right].$$

The between imputation variance is given by, $\hat{B} = \frac{1}{K-1} \sum_{k=1}^K (\hat{\theta}_{\text{CR, } k} - \hat{\theta}_{\text{CR, MI}})^2$. Taking the necessary expectations, as done in Subsection 3.2.1 reveals,

$$E[\hat{B}] = \sigma_{2.1} \pi_d \frac{(\pi_d + 1)}{(n-1)}.$$

Thus the overall expected value of the variance of the MI estimator under CR, where we assume that n is sufficiently large so that we may take $(n - 1)$ to be n is,

$$E[V_{\text{CR, MI}}] = \frac{2\sigma_{2.1}}{n} + \frac{\Delta^2\pi_d(1 - \pi_d)}{n} + \frac{2\sigma_{2.1}\pi_d}{n^2} \left[2 - \pi_d \left(1 + \frac{1}{n} \right) \right] + \left(1 + \frac{1}{K} \right) \sigma_{2.1}\pi_d \frac{(\pi_d + 1)}{n}.$$

For infinite K this is equal to,

$$E[V_{\text{CR, MI}}] = \frac{2\sigma_{2.1}}{n} + \frac{\Delta^2\pi_d(1 - \pi_d)}{n} + \frac{2\sigma_{2.1}\pi_d}{n^2} \left[2 - \pi_d \left(1 + \frac{1}{n} \right) \right] + \sigma_{2.1}\pi_d \frac{(\pi_d + 1)}{n}.$$

Table B.4 summarises Rubin's derived variance estimator under CR alongside simulation results. Derived results are within 2 MCSE's of the simulation results. The expectation of the ideal information anchored variance under CR is,

$$E[V_{\text{anchored}}] = \frac{2\sigma_{2.1}}{n} + \frac{\Delta^2\pi_d(1 - \pi_d)}{n} + \frac{\Delta^2\pi_d^2}{2n} + \frac{\sigma_{2.1}\pi_d}{(1 - \pi_d)n},$$

which up to the π_d^4/n terms, is equivalent to,

$$E[V_{\text{anchored}}] = \frac{2\sigma_{2.1}}{n} + \frac{\Delta^2\pi_d(1 - \pi_d)}{n} + \frac{\Delta^2\pi_d^2}{2n} + \frac{\sigma_{2.1}\pi_d}{n} + \frac{\sigma_{2.1}\pi_d^2}{n} + \frac{\sigma_{2.1}}{n} [\pi_d^3 + \pi_d^4].$$

The difference between Rubin's MI variance estimator under CR where $K \rightarrow \infty$ and the ideal information anchored variance is,

$$E[V_{\text{CR, MI}}] - E[V_{\text{anchored}}] = \frac{2\sigma_{2.1}\pi_d}{n^2} \left[2 - \pi_d \left(1 + \frac{1}{n} \right) \right] - \frac{\Delta^2\pi_d^2}{2n} - \frac{\sigma_{2.1}\pi_d^3}{n} - \frac{\sigma_{2.1}\pi_d^4}{n}.$$

The largest terms of the difference are,

$$-\frac{\Delta^2\pi_d^2}{2n} - \frac{\sigma_{2.1}\pi_d^3}{n} - \frac{\sigma_{2.1}\pi_d^4}{n}. \quad (3.42)$$

We see that the difference is very small and approaches zero for larger n . The difference is not the same as in the unadjusted setting. But similar to the unadjusted setting discussed in Subsection

3.2.1, the precise difference depends on the size of the trial (n), the amount of missingness in the active arm (n_d), the correlation between the time 1 and time 2 measurement, the variance of the data at time 2 ($\sigma_{2.1} = \sigma_{22}(1 - \rho^2)$), as well as the mean difference between the active and reference arm at time 2 (Δ). For realistic trial settings (powered at 80% with 5% statistical significance) where $\Delta^2 \approx 15.68\sigma_{22}/n$ the difference is at most $O(\pi_d^3 n^{-1})$. The first term in (3.42) will be $O(n^{-2})$ and the second two terms are respectively $O(\pi_d^3 n^{-1})$ and $O(\pi_d^4 n^{-1})$. The information anchored variance is of a larger order, $O(n^{-1})$. The terms in (3.42) are therefore practically negligible in comparison. Under CR the ratio of the difference between Rubin's variance estimator and the information anchored variance to the information anchored variance is,

$$\frac{\pi_d^2 [-\Delta^2 - 2\sigma_{22}(1 - \rho^2)(\pi_d + \pi_d^2)]}{4\sigma_{2.1} + \pi_d [\Delta^2 (2(1 - \pi_d) + \pi_d) + 2\sigma_{22}(1 - \rho^2)(1 + \pi_d + \pi_d^2 + \pi_d^3)]}.$$

Thus Rubin's variance estimator provides an excellent approximation for the information anchored variance estimate that preserves the information loss in the primary analysis in the baseline adjusted CR setting.

Rubin's variance estimator under J2R

We follow the same set-up and initial developments for an appropriate imputation distribution under J2R as in the unadjusted treatment effect setting outlined in Subsection 3.2.2. The imputation model is as in (3.18).

Under the described conditions the baseline adjusted treatment estimator in imputation k is computed as, $\hat{\theta}_{\text{J2R}, k} = \bar{Y}_{a2,k} - \bar{Y}_{r2}$. The expectation of $\bar{Y}_{a2,k}$ and \bar{Y}_{r2} have already been computed under J2R in Subsection 3.2.2, which reveals $E[\hat{\theta}_{\text{J2R}, k}] = (n_o/n)(\mu_{a2} - \mu_{r2})$, thus $E[\hat{\theta}_{\text{J2R}, \text{MI}}] = (n_o/n)(\mu_{a2} - \mu_{r2})$.

For Rubin's variance estimator we require,

$$E[\hat{W}] = \frac{\sigma_{2.1}}{n} + \frac{E[\hat{\sigma}_{2.1,a,k}]}{n},$$

where,

$$E[\hat{\sigma}_{2.1,a,k}] = E[\hat{\sigma}_{22,a,k}] - \frac{\sigma_{12}^2}{\sigma_{11}}.$$

The expectation of $\hat{\sigma}_{22,a,k}$ under J2R as computed in Subsection 3.2.2 is,

$$E[\hat{\sigma}_{22,a,k}] = \sigma_{22} + \frac{\Delta^2 \pi_d n_o}{(n-1)} + \frac{\sigma_{22} \pi_d (2 - \pi_d)}{(n-1)} + \frac{\sigma_{12}^2 \pi_d (2 - 3\pi_d)}{\sigma_{11} (n-1)} \\ + \frac{\sigma_{2.1} \pi_d}{(n-1)} \left[\frac{2n}{(n-1)} - \frac{2\pi_d}{(n-1)} - \pi_d \right]$$

where $\Delta = (\mu_{a2} - \mu_{r2})$. Taking the necessary expectations, as done in Subsection 3.2.2, reveals the between imputation variance is,

$$E[\hat{B}] = \pi_d^2 \left[\frac{\sigma_{2.1}}{(n-1)} \left(\frac{2}{n} + 1 + \frac{1}{\pi_d} \right) + \frac{2 \sigma_{12}^2}{n \sigma_{11}} \right].$$

Thus the overall expected value of the variance of the MI estimator under J2R, where we assume that n is sufficiently large so that we may take $(n-1)$ to be n is,

$$E[V_{J2R, MI}] = \frac{2\sigma_{2.1}}{n} + \frac{\Delta^2 \pi_d (1 - \pi_d)}{n} + \frac{\sigma_{22} \pi_d (2 - \pi_d)}{n^2} + \frac{\sigma_{12}^2 \pi_d (2 - 3\pi_d)}{\sigma_{11} n^2} \\ + \frac{\sigma_{2.1} \pi_d}{n^2} \left(2 - \frac{2\pi_d}{n} - \pi_d \right) + \left(1 + \frac{1}{K} \right) \frac{\pi_d^2}{n} \left[\sigma_{2.1} \left(\frac{2}{n} + 1 + \frac{1}{\pi_d} \right) + 2 \frac{\sigma_{12}^2}{\sigma_{11}} \right].$$

For infinite K this is equal to,

$$E[V_{J2R, MI}] = \frac{2\sigma_{2.1}}{n} + \frac{\Delta^2 \pi_d (1 - \pi_d)}{n} + \frac{\sigma_{2.1} \pi_d}{n} + \frac{\sigma_{2.1} \pi_d^2}{n} + 2 \frac{\sigma_{12}^2 \pi_d^2}{\sigma_{11} n} + \frac{\sigma_{12}^2 \pi_d (2 - 3\pi_d)}{\sigma_{11} n^2} \\ + \frac{\sigma_{2.1} \pi_d}{n^2} \left[2 + \pi_d \left(1 - \frac{2}{n} \right) \right] + \frac{\sigma_{22} \pi_d (2 - \pi_d)}{n^2}.$$

Table B.6 shows Rubin's derived variance estimator under J2R is within 2 MCSE's of simulation results. The expectation of the ideal information anchored variance under J2R up to the order π_d^4/n terms, where $\Delta = (\mu_{a2} - \mu_{r2})$ is,

$$E[V_{\text{anchored}}] = \frac{2\sigma_{2.1}}{n} + \frac{\Delta^2 \pi_d (1 - \pi_d)}{n} + \frac{\Delta^2 \pi_d^2}{2n} + \frac{\sigma_{2.1} \pi_d}{n} + \frac{\sigma_{2.1} \pi_d^2}{n} + \frac{\sigma_{2.1}}{n} [\pi_d^3 + \pi_d^4].$$

The difference between Rubin's MI variance estimator where $K \rightarrow \infty$ and the ideal information anchored variance is,

$$E[V_{\text{J2R, MI}}] - E[V_{\text{anchored}}] = \frac{\sigma_{12}^2 \pi_d}{\sigma_{11} n} \left(2\pi_d + \frac{2 - 3\pi_d}{n} \right) + \frac{\sigma_{2.1}\pi_d}{n^2} \left[2 + \frac{\pi_d(n-2)}{n} \right] \\ + \frac{\sigma_{22}\pi_d(2 - \pi_d)}{n^2} - \frac{\Delta^2\pi_d^2}{2n} - \frac{\sigma_{2.1}\pi_d^3}{n} - \frac{\sigma_{2.1}\pi_d^4}{n}.$$

The terms with the largest magnitude in the difference are,

$$\frac{2\pi_d^2 \sigma_{12}^2}{n \sigma_{11}} - \frac{\Delta^2\pi_d^2}{2n} - \frac{\sigma_{2.1}\pi_d^3}{n} - \frac{\sigma_{2.1}\pi_d^4}{n}. \quad (3.43)$$

The difference is very small and approaches zero for larger n . The difference again depends on the size of the trial (n), the amount of missingness in the active arm (n_d), the correlation between the time 1 and time 2 measurement, the variance of the data at time 2 ($\sigma_{2.1} = \sigma_{22}(1 - \rho^2)$), as well as the mean difference between the active and reference arm at time 2 (Δ). But for realistic trial settings the difference is tiny. The first term in (3.43) will be $O(\pi_d^2 n^{-1})$ therefore practically negligible in comparison to the information anchored variance we desire, which is $O(n^{-1})$. The remaining three terms in the difference are of an even smaller order than that of the first term in (3.43). Under J2R the ratio of the difference between Rubin's variance estimator and the information anchored variance to the information anchored variance is,

$$\frac{\pi_d^2 [-\Delta^2 - 2\sigma_{22}(1 - \rho^2)(\pi_d + \pi_d^2)] + 4\sigma_{12}^2\sigma_{11}^{-1}\pi_d^2}{4\sigma_{2.1} + \pi_d[\Delta^2(2(1 - \pi_d) + \pi_d) + 2\sigma_{22}(1 - \rho^2)(1 + \pi_d + \pi_d^2 + \pi_d^3)]}.$$

Thus Rubin's variance estimator also well approximates the ideal information anchored variance in the J2R baseline adjusted setting.

Rubin's variance estimator under copy increments in reference

We follow the same set-up and initial developments for an appropriate imputation distribution under CIR as in the unadjusted treatment effect setting outlined in Subsection 3.2.3. The imputation model is as in (3.25).

The baseline adjusted treatment estimator in imputation k is computed as, $\hat{\theta}_{\text{CIR}, k} = \bar{Y}_{a2,k} - \bar{Y}_{r2}$. The expectation of $\bar{Y}_{a2,k}$ and \bar{Y}_{r2} have already been computed under CIR in Subsection 3.2.3, which reveals $E[\hat{\theta}_{\text{CIR}, k}] = (n_o/n)(\mu_{a2} - \mu_{r2})$ and $E[\hat{\theta}_{\text{CIR, MI}}] = (n_o/n)(\mu_{a2} - \mu_{r2})$.

For Rubin's variance estimator we require,

$$E[\hat{W}] = \frac{\sigma_{2.1}}{n} + \frac{E[\hat{\sigma}_{2.1,a,k}]}{n},$$

where,

$$E[\hat{\sigma}_{2.1,a,k}] = E[\hat{\sigma}_{22,a,k}] - \frac{\sigma_{12}^2}{\sigma_{11}}.$$

The expectation of $\hat{\sigma}_{22,a,k}$ under CIR, as computed in Subsection 3.2.3 is,

$$\begin{aligned} E[\hat{\sigma}_{22,a,k}] = & \sigma_{22} + \frac{\Delta^2 \pi_d n_o}{(n-1)} + \frac{\sigma_{22} \pi_d (2 - \pi_d)}{(n-1)} + \frac{\sigma_{12}^2 \pi_d (2 - 3\pi_d)}{\sigma_{11} (n-1)} \\ & + \frac{\sigma_{2.1} \pi_d}{(n-1)} \left[\frac{2n}{(n-1)} - \frac{2\pi_d}{(n-1)} - \pi_d \right] + \frac{4\pi_d (1 - \pi_d)}{(n-1)} (\sigma_{11} - 2\sigma_{12}), \end{aligned}$$

for $\Delta = (\mu_{a2} - \mu_{r2})$. Taking the necessary expectations reveals the between imputation variance is,

$$E[\hat{B}] = \pi_d^2 \left[\frac{\sigma_{2.1}}{(n-1)} \left(\frac{2}{n} + 1 + \frac{1}{\pi_d} \right) + \frac{2}{n} \left(\frac{\sigma_{12}^2}{\sigma_{11}} + \sigma_{11} - 2\sigma_{12} \right) \right].$$

Thus the overall expected value of the variance of the MI estimator under CIR, where we assume that n is sufficiently large so that we may take $(n-1)$ to be n is,

$$\begin{aligned} E[V_{\text{CIR, MI}}] = & \frac{2\sigma_{2.1}}{n} + \frac{\Delta^2 \pi_d (1 - \pi_d)}{n} + \frac{\sigma_{22} \pi_d (2 - \pi_d)}{n^2} + \frac{\sigma_{12}^2 \pi_d (2 - 3\pi_d)}{\sigma_{11} n^2} \\ & + \frac{\sigma_{2.1} \pi_d}{n^2} \left[2 - \frac{2\pi_d}{n} - \pi_d \right] + \frac{4\pi_d (1 - \pi_d)}{n^2} (\sigma_{11} - 2\sigma_{12}) \\ & + \left(1 + \frac{1}{K} \right) \frac{\pi_d^2}{n} \left[\sigma_{2.1} \left(\frac{2}{n} + 1 + \frac{1}{\pi_d} \right) + 2 \left(\frac{\sigma_{12}^2}{\sigma_{11}} + \sigma_{11} - 2\sigma_{12} \right) \right]. \end{aligned}$$

For infinite K this is,

$$\begin{aligned} E[V_{\text{CIR, MI}}] = & \frac{2\sigma_{2.1}}{n} + \frac{\Delta^2 \pi_d (1 - \pi_d)}{n} + \frac{\sigma_{2.1} \pi_d}{n} + \frac{\sigma_{2.1} \pi_d^2}{n} + 2 \frac{\sigma_{12}^2 \pi_d^2}{\sigma_{11} n} + \frac{\sigma_{12}^2 \pi_d (2 - 3\pi_d)}{\sigma_{11} n^2} \\ & + \frac{\sigma_{2.1} \pi_d}{n^2} \left[2 + \pi_d \left(1 - \frac{2}{n} \right) \right] + \frac{\sigma_{22} \pi_d (2 - \pi_d)}{n^2} \\ & + 2\pi_d (\sigma_{11} - 2\sigma_{12}) \left[\frac{(n-2)\pi_d + 2}{n^2} \right]. \end{aligned}$$

Table B.8 shows Rubin's derived variance estimator under CIR is within 2 MCSE's of simulation results. The expectation of the ideal information anchored variance under CIR up to the π_d^4/n terms, where $\Delta = (\mu_{a2} - \mu_{r2})$ is,

$$E[V_{\text{anchored}}] = \frac{2\sigma_{2.1}}{n} + \frac{\Delta^2 \pi_d (1 - \pi_d)}{n} + \frac{\Delta^2 \pi_d^2}{2n} + \frac{\sigma_{2.1} \pi_d}{n} + \frac{\sigma_{2.1} \pi_d^2}{n} + \frac{\sigma_{2.1}}{n} [\pi_d^3 + \pi_d^4].$$

The difference between Rubin's MI variance estimator where $K \rightarrow \infty$ and the ideal information anchored variance is,

$$\begin{aligned} E[V_{\text{anchored}}] - E[V_{\text{CIR, MI}}] &= \frac{\sigma_{12}^2 \pi_d}{\sigma_{11} n} \left(2\pi_d + \frac{2 - 3\pi_d}{n} \right) + \frac{\sigma_{2.1} \pi_d}{n^2} \left[2 + \frac{\pi_d (n - 2)}{n} \right] \\ &\quad + \frac{\sigma_{22} \pi_d (2 - \pi_d)}{n^2} + 2\pi_d (\sigma_{11} - 2\sigma_{12}) \left[\frac{(n - 2) \pi_d + 2}{n^2} \right] \\ &\quad - \frac{\Delta^2 \pi_d^2}{2n} - \frac{\sigma_{2.1} \pi_d^3}{n} - \frac{\sigma_{2.1} \pi_d^4}{n}. \end{aligned}$$

The difference is dominated by,

$$\frac{\sigma_{12}^2 \pi_d^2}{\sigma_{11} n} + \frac{2\pi_d^2 (\sigma_{11} - 2\sigma_{12})}{n} - \frac{\Delta^2 \pi_d^2}{2n} - \frac{\sigma_{2.1} \pi_d^3}{n} - \frac{\sigma_{2.1} \pi_d^4}{n}.$$

The difference is again very small and approaches zero for larger n . As in the unadjusted setting the difference here matches the difference seen under J2R, plus an additional component, which is also very small and $O(\pi_d^2 n^{-1})$. This will again be practically negligible in comparison to the information anchored variance we desire, which is $O(n^{-1})$. Under CIR the ratio of the difference between Rubin's variance estimator and the information anchored variance to the information anchored variance is,

$$\frac{\pi_d^2 [-\Delta^2 - 2\sigma_{22} (1 - \rho^2) (\pi_d + \pi_d^2)] + 4\sigma_{12}^2 \sigma_{11}^{-1} \pi_d^2 + 4(\sigma_{11} - 2\sigma_{12}) \pi_d^2}{4\sigma_{2.1} + \pi_d [\Delta^2 (2(1 - \pi_d) + \pi_d) + 2\sigma_{22} (1 - \rho^2) (1 + \pi_d + \pi_d^2 + \pi_d^3)]}.$$

Thus we see Rubin's variance estimator also approximates the ideal information anchored variance under CIR for the baseline adjusted treatment effect.

Rubin's variance estimator under last mean carried forward

Finally for imputation under LMCF we follow the same set-up and initial developments for an appropriate imputation distribution as in the unadjusted treatment effect setting outlined in Subsection 3.2.4. The imputation model is as in (3.32).

The baseline adjusted treatment estimator in imputation k is computed as, $\hat{\theta}_{\text{LMCF}, k} = \bar{Y}_{a2, k} - \bar{Y}_{r2}$. The expectation of $\bar{Y}_{a2, k}$ and \bar{Y}_{r2} have already been computed under LMCF in Subsection 3.2.3, which reveals $E[\hat{\theta}_{\text{LMCF}, k}] = (n_o/n)\mu_{a2} + (n_d/n)\mu_{a1} - \mu_{r2}$ and $E[\hat{\theta}_{\text{LMCF}, \text{MI}}] = (n_o/n)\mu_{a2} +$

$$(n_d/n)\mu_{a1} - \mu_{r2}.$$

For Rubin's variance estimator we require,

$$E[\hat{W}] = \frac{\sigma_{2.1}}{n} + \frac{E[\hat{\sigma}_{2.1,a,k}]}{n},$$

where,

$$E[\hat{\sigma}_{2.1,a,k}] = E[\hat{\sigma}_{22,a,k}] - \frac{\sigma_{12}^2}{\sigma_{11}}.$$

The expectation of $\hat{\sigma}_{22,a,k}$ under LMCF, as computed in Subsection 3.2.4 is,

$$E[\hat{\sigma}_{22,a,k}] = \sigma_{22} + \frac{\Delta^2 \pi_d n_o}{(n-1)} + \frac{2\pi_d(1-\pi_d)}{(n-1)} \left[\sigma_{11} - 2\sigma_{12} + \frac{\sigma_{12}^2}{\sigma_{11}} \right] + \frac{2\sigma_{2.1}\pi_d}{(n_o-1)},$$

for $\Delta = (\mu_{a2} - \mu_{a1})$. Taking the necessary expectations as in Subsection 3.2.4 reveals the between imputation variance where we assume that n is sufficiently large so that we may take $(n_o - 1)$ to be n_o is,

$$E[\hat{B}] = \frac{\pi_d^2}{n} \left[\sigma_{11} - 2\sigma_{12} + \frac{\sigma_{12}^2}{\sigma_{11}} + \sigma_{2.1} \left(\frac{1}{\pi_d} + \frac{1}{n(1-\pi_d)} \right) \right].$$

Thus the overall expected value of the variance of the MI estimator under LMCF, where we assume that n is sufficiently large so that we may take $(n-1)$ to be n and (n_o-1) to be n_o is,

$$\begin{aligned} E[V_{\text{LMCF, MI}}] &= \frac{2\sigma_{2.1}}{n} + \frac{\Delta^2 \pi_d(1-\pi_d)}{n} + \frac{2\pi_d(1-\pi_d)}{n^2} \left[\sigma_{11} - 2\sigma_{12} + \frac{\sigma_{12}^2}{\sigma_{11}} \right] + \frac{2\sigma_{2.1}\pi_d}{n^2(1-\pi_d)} \\ &\quad + \left(1 + \frac{1}{K} \right) \frac{\pi_d^2}{n} \left[\sigma_{11} - 2\sigma_{12} + \frac{\sigma_{12}^2}{\sigma_{11}} + \sigma_{2.1} \left(\frac{1}{\pi_d} + \frac{1}{n(1-\pi_d)} \right) \right]. \end{aligned}$$

For infinite K this is,

$$\begin{aligned} E[V_{\text{LMCF, MI}}] &= \frac{2\sigma_{2.1}}{n} + \frac{\Delta^2 \pi_d(1-\pi_d)}{n} + \frac{\sigma_{2.1}\pi_d}{n} \left[1 + \frac{2}{n(1-\pi_d)} + \frac{\pi_d}{n(1-\pi_d)} \right] \\ &\quad + \frac{\pi_d(2+(n-2)\pi_d)}{n^2} \left[\sigma_{11} - 2\sigma_{12} + \frac{\sigma_{12}^2}{\sigma_{11}} \right]. \end{aligned}$$

The expectation of the ideal information anchored variance under LMCF up to the π_d^4/n terms, where $\Delta = (\mu_{a2} - \mu_{a1})$ is,

$$E[V_{\text{anchored}}] = \frac{2\sigma_{2.1}}{n} + \frac{\Delta^2 \pi_d (1 - \pi_d)}{n} + \frac{\Delta^2 \pi_d^2}{2n} + \frac{\sigma_{2.1} \pi_d}{n} + \frac{\sigma_{2.1} \pi_d^2}{n} + \frac{\sigma_{2.1}}{n} [\pi_d^3 + \pi_d^4].$$

Table B.10 summarises Rubin's derived variance estimator under LMCF alongside simulation results. Derived results are within 2 MCSE's of the simulation results. The difference between Rubin's MI variance estimator where $K \rightarrow \infty$ and the ideal information anchored variance is,

$$\begin{aligned} E[V_{\text{anchored}}] - E[V_{\text{LMCF, MI}}] &= \frac{\sigma_{2.1} \pi_d}{n} \left[\frac{2}{n(1 - \pi_d)} + \frac{\pi_d}{n(1 - \pi_d)} \right] \\ &+ \frac{\pi_d (2 + (n - 2) \pi_d)}{n^2} \left[\sigma_{11} - 2\sigma_{12} + \frac{\sigma_{12}^2}{\sigma_{11}} \right] - \frac{\Delta^2 \pi_d^2}{2n} \\ &- \frac{\sigma_{2.1} \pi_d^2}{n} - \frac{\sigma_{2.1} \pi_d^3}{n} - \frac{\sigma_{2.1} \pi_d^4}{n}. \end{aligned}$$

The difference is dominated by,

$$\frac{\pi_d^2}{n} \left[\sigma_{11} - 2\sigma_{12} + \frac{\sigma_{12}^2}{\sigma_{11}} \right] - \frac{\Delta^2 \pi_d^2}{2n} - \frac{\sigma_{2.1} \pi_d^2}{n} - \frac{\sigma_{2.1} \pi_d^3}{n} - \frac{\sigma_{2.1} \pi_d^4}{n}. \quad (3.44)$$

Again the difference is very small and approaches zero for larger n . Similar to the unadjusted setting, the precise difference depends on the size of the trial (n), the amount of missingness in the active arm (n_d), the correlation between the time 1 and time 2 measurement, the variance of the data at time 2 ($\sigma_{2.1} = \sigma_{22}(1 - \rho^2)$), as well as the mean difference between the active arm mean at time 1 and time 2 (Δ). The largest term in (3.44) is $O(\pi_d^2 n^{-1})$. This is tiny in comparison to the information anchored variance, which is $O(n^{-1})$. Under LMCF the ratio of the difference between Rubin's variance estimator and the information anchored variance to the information anchored variance is,

$$\frac{\pi_d^2 [-\Delta^2 - 2\sigma_{22} (1 - \rho^2) (1 + \pi_d + \pi_d^2)] + 2\pi_d^2 (\sigma_{11} - 2\sigma_{12} + \sigma_{12}^2 \sigma_{11}^{-1})}{4\sigma_{2.1} + \pi_d [\Delta^2 (2(1 - \pi_d) + \pi_d) + 2\sigma_{22} (1 - \rho^2) (1 + \pi_d + \pi_d^2 + \pi_d^3)]}.$$

We see that Rubin's variance estimator also provides an excellent approximation for the ideal information anchored variance under LMCF.

We see in all the reference based settings the difference between Rubin's variance estimator and the ideal information anchored variance estimate for the baseline adjusted treatment effect is of a very small order. Rubin's variance estimator approximates the desired information anchored variance for the adjusted treatment effect, thus keeps the information loss constant across primary and sensitivity analysis.

The results for Rubin's variance estimator for the baseline adjusted treatment effect are similar to those for the unadjusted treatment effect. The inference proceeds as expected. The difference is due to replacing a marginal variance with a conditional variance. However this difference is minimal. In Table 3.4 we summarise the difference between Rubin's MI variance estimator and the information anchored variance for each imputation procedure for both the unadjusted and baseline treatment estimator. For the baseline adjusted treatment estimators we have replaced $\sigma_{2.1} = \sigma_{22}(1 - \rho^2)$ for comparison.

Imputation Method	Rubin's MI variance - Information anchored variance (largest order terms)
Unadjusted treatment estimator:	
CR	$-\frac{\Delta^2(1-\rho^2)\pi_d^2}{2n} - \frac{\sigma_{22}(1-\rho^2)\pi_d^3}{n} - \frac{\sigma_{22}(1-\rho^2)\pi_d^4}{n}$
J2R	$-\frac{\Delta^2(1-\rho^2)\pi_d^2}{2n} - \frac{\sigma_{22}(1-\rho^2)\pi_d^3}{n} - \frac{\sigma_{22}(1-\rho^2)\pi_d^4}{n} + 2\frac{\sigma_{12}^2}{\sigma_{11}}\frac{\pi_d^2}{n}$
CIR	$-\frac{\Delta^2(1-\rho^2)\pi_d^2}{2n} - \frac{\sigma_{22}(1-\rho^2)\pi_d^3}{n} - \frac{\sigma_{22}(1-\rho^2)\pi_d^4}{n} + 2\frac{\sigma_{12}^2}{\sigma_{11}}\frac{\pi_d^2}{n} + 2(\sigma_{11} - 2\sigma_{12})\frac{\pi_d^2}{n}$
LMCF	$-\frac{\Delta^2(1-\rho^2)\pi_d^2}{2n} - \frac{\sigma_{22}(1-\rho^2)\pi_d^3}{n} - \frac{\sigma_{22}(1-\rho^2)\pi_d^4}{n} - \frac{\sigma_{22}(1-\rho^2)\pi_d^4}{n} + \frac{\pi_d^2}{n} \left[\sigma_{11} - 2\sigma_{12} + \frac{\sigma_{12}^2}{\sigma_{11}} \right]$
Baseline adjusted treatment estimator:	
CR	$-\frac{\Delta^2\pi_d^2}{2n} - \frac{\sigma_{22}(1-\rho^2)\pi_d^3}{n} - \frac{\sigma_{22}(1-\rho^2)\pi_d^4}{n}$
J2R	$-\frac{\Delta^2\pi_d^2}{2n} - \frac{\sigma_{22}(1-\rho^2)\pi_d^3}{n} - \frac{\sigma_{22}(1-\rho^2)\pi_d^4}{n} + 2\frac{\sigma_{12}^2}{\sigma_{11}}\frac{\pi_d^2}{n}$
CIR	$-\frac{\Delta^2\pi_d^2}{2n} - \frac{\sigma_{22}(1-\rho^2)\pi_d^3}{n} - \frac{\sigma_{22}(1-\rho^2)\pi_d^4}{n} + 2\frac{\sigma_{12}^2}{\sigma_{11}}\frac{\pi_d^2}{n} + 2(\sigma_{11} - 2\sigma_{12})\frac{\pi_d^2}{n}$
LMCF	$-\frac{\Delta^2\pi_d^2}{2n} - \frac{\sigma_{22}(1-\rho^2)\pi_d^3}{n} - \frac{\sigma_{22}(1-\rho^2)\pi_d^4}{n} - \frac{\sigma_{22}(1-\rho^2)\pi_d^4}{n} + \frac{\pi_d^2}{n} \left[\sigma_{11} - 2\sigma_{12} + \frac{\sigma_{12}^2}{\sigma_{11}} \right]$

Table 3.4: Difference between Rubin's MI variance estimator and the information anchored variance for CR, J2R, CIR and LMCF imputation with baseline and a single follow-up and baseline adjustment. Note: $\Delta = \mu_{a2} - \mu_{r2}$ for CR, J2R and CIR. For LMCF $\Delta = \mu_{a2} - \mu_{a1}$.

We see the only difference is in the first term of each equation, which involves the Δ term. There is an additional multiplier of $(1 - \rho^2)$ in the unadjusted case due to the marginal variance. However the overall magnitude of this component in both cases is tiny, that is $O(n^{-2})$. Thus the additional multiplier will not have a practically important impact. In Table 3.5 we summarise the ratio of the difference between Rubin's MI variance estimator and the information anchored variance to the information anchored variance for each imputation procedure for both the unadjusted and baseline treatment estimator. The baseline adjustment does not greatly effect the information anchoring properties of Rubin's variance estimator. We now demonstrate this via a simulation study.

Imputation Method	Rubin's MI variance - Information anchored variance (largest order terms)
Unadjusted treatment estimator:	
CR	$\frac{(1-\rho^2)\pi_d^2[-\Delta^2-2\sigma_{22}(\pi_d+\pi_d^2)]}{4\sigma_{22}+\pi_d[\Delta^2(2(1-\pi_d)+(1-\rho^2)\pi_d)+2\sigma_{22}(1-\rho^2)(1+\pi_d+\pi_d^2+\pi_d^3)]}$
J2R	$\frac{(1-\rho^2)\pi_d^2[-\Delta^2-2\sigma_{22}(\pi_d+\pi_d^2)]+4\sigma_{12}^2\sigma_{11}^{-1}\pi_d^2}{4\sigma_{22}+\pi_d[\Delta^2(2(1-\pi_d)+(1-\rho^2)\pi_d)+2\sigma_{22}(1-\rho^2)(1+\pi_d+\pi_d^2+\pi_d^3)]}$
CIR	$\frac{(1-\rho^2)\pi_d^2[-\Delta^2-2\sigma_{22}(\pi_d+\pi_d^2)]+4\sigma_{12}^2\sigma_{11}^{-1}\pi_d^2+4\pi_d(\sigma_{11}-2\sigma_{12})\pi_d}{4\sigma_{22}+\pi_d[\Delta^2(2(1-\pi_d)+(1-\rho^2)\pi_d)+2\sigma_{22}(1-\rho^2)(1+\pi_d+\pi_d^2+\pi_d^3)]}$
LMCF	$\frac{(1-\rho^2)\pi_d^2[-\Delta^2-2\sigma_{22}(1+\pi_d+\pi_d^2)]+2\pi_d^2(\sigma_{11}-2\sigma_{12}+\sigma_{12}^2\sigma_{11}^{-1})}{4\sigma_{22}+\pi_d[\Delta^2(2(1-\pi_d)+(1-\rho^2)\pi_d)+2\sigma_{22}(1-\rho^2)(1+\pi_d+\pi_d^2+\pi_d^3)]}$
Baseline adjusted treatment estimator:	
CR	$\frac{\pi_d^2[-\Delta^2-2\sigma_{22}(1-\rho^2)(\pi_d+\pi_d^2)]}{4\sigma_{2.1}+\pi_d[\Delta^2(2(1-\pi_d)+\pi_d)+2\sigma_{22}(1-\rho^2)(1+\pi_d+\pi_d^2+\pi_d^3)]}$
J2R	$\frac{\pi_d^2[-\Delta^2-2\sigma_{22}(1-\rho^2)(\pi_d+\pi_d^2)]+4\sigma_{12}^2\sigma_{11}^{-1}\pi_d^2}{4\sigma_{2.1}+\pi_d[\Delta^2(2(1-\pi_d)+\pi_d)+2\sigma_{22}(1-\rho^2)(1+\pi_d+\pi_d^2+\pi_d^3)]}$
CIR	$\frac{\pi_d^2[-\Delta^2-2\sigma_{22}(1-\rho^2)(\pi_d+\pi_d^2)]+4\sigma_{12}^2\sigma_{11}^{-1}\pi_d^2+4(\sigma_{11}-2\sigma_{12})\pi_d^2}{4\sigma_{2.1}+\pi_d[\Delta^2(2(1-\pi_d)+\pi_d)+2\sigma_{22}(1-\rho^2)(1+\pi_d+\pi_d^2+\pi_d^3)]}$
LMCF	$\frac{\pi_d^2[-\Delta^2-2\sigma_{22}(1-\rho^2)(1+\pi_d+\pi_d^2)]+2\pi_d^2(\sigma_{11}-2\sigma_{12}+\sigma_{12}^2\sigma_{11}^{-1})}{4\sigma_{2.1}+\pi_d[\Delta^2(2(1-\pi_d)+\pi_d)+2\sigma_{22}(1-\rho^2)(1+\pi_d+\pi_d^2+\pi_d^3)]}$

Table 3.5: The ratio of the difference between Rubin's MI variance estimate and the information anchored variance to the information anchored variance for CR, J2R, CIR and LMCF imputation with baseline and a single follow-up and baseline adjustment. Note: $\Delta = \mu_{a2} - \mu_{r2}$ for CR, J2R and CIR. For LMCF $\Delta = \mu_{a2} - \mu_{a1}$.

3.4.3 Simulation study to investigate Rubin’s variance estimator for the baseline adjusted treatment estimator

We present the results of simulation study conducted to evaluate the information anchoring properties of Rubin’s MI variance estimator in a variety of trial scenarios, for the baseline adjusted treatment effect in each of the reference based settings (CR, J2R, CIR and LMCF). The baseline and week 12 data from the asthma trial described in Subsection 1.7.2 formed the motivating example for generating the datasets.

Methods

Baseline (time 1) and follow-up (time 2) data were generated using an underlying bivariate normal distribution with means and covariance matrices similar to those obtained in the asthma study at baseline and week 12 as,

$$\mu_{\text{placebo}} = [2, 1.9], \mu_{\text{active}} = [2, \mu_{a,2}],$$

$$\Sigma_{\text{placebo}} = \Sigma_{\text{active}} = \begin{bmatrix} 0.4 & 0.2 \\ 0.2 & 0.6 \end{bmatrix},$$

with a range of $\mu_{a,2}$ corresponding to treatment effects of $\Delta = \{0, 0.3, 1\}$ and a sample size of $n=250$ per arm. Interest now lies in the baseline adjusted mean treatment group difference at time 2. The analysis model was a linear regression of the time 2 response on randomised treatment group, with adjustment for baseline. For each setting (unique Δ), the treatment effect and estimated variance were computed using the full data for 1000 simulations. After imposing the missingness mechanisms outlined in Subsection 3.3.1, corresponding to 10–70% missingness at time 2 in the active arm, the reference based MI approach of Carpenter, Roger and Kenward [13] was then applied to each incomplete dataset under MAR, CR, J2R, CIR and LMCF with 50 imputations. In each scenario post-deviation data were regenerated using the appropriate conditional normal distributions to compute the baseline adjusted treatment effect and its variance had the post-deviation data been observed in the particular de-facto scenario.

The main outcomes of interest were Rubin’s MI estimate of variance and the ideal information anchored variance computed as in Subsection 3.4.1 for the baseline adjusted treatment effect. Estimates were averaged over the 1000 simulations in each scenario for comparison.

Results

In simulations with a treatment effect of $\Delta = 0.3$ ($\mu_{a,2} = 2.2$), as observed in the original asthma trial, excellent agreement is seen between Rubin’s MI variance estimator and the information anchored variance for up to 40–50% missing data in the active arm. This corresponds with the results seen for the unadjusted treatment effect. Excellent agreement up to 40–50% missingness is also observed with smaller and larger treatment effects (see Figures C.11 and C.12).

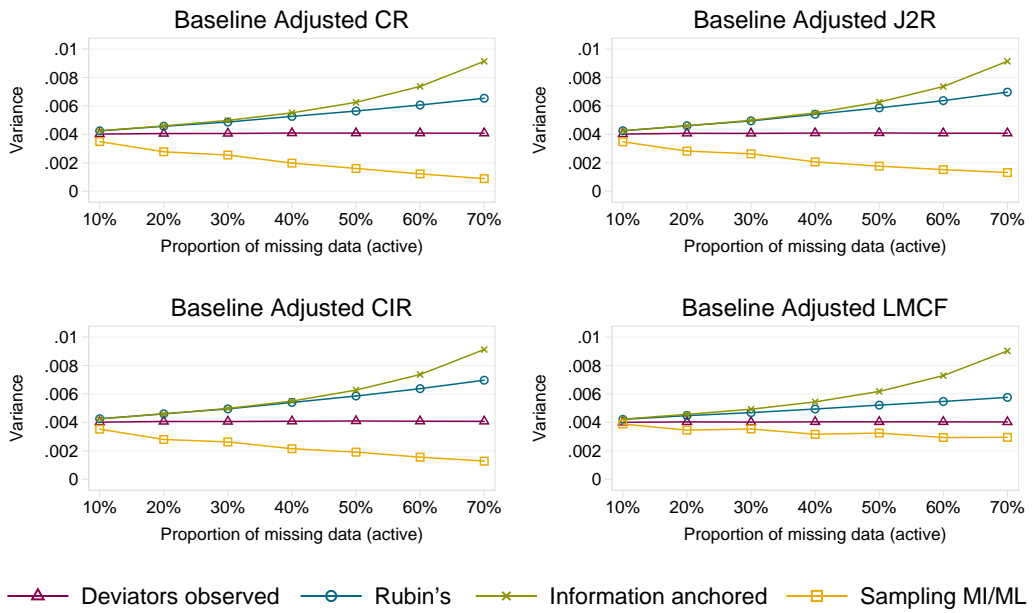


Figure 3.6: Rubin's baseline adjusted variance estimator vs. the baseline adjusted information anchored variance vs. baseline adjusted variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 baseline adjusted MI estimates, with the asthma trial treatment effect of $\Delta = 0.3$ and $n=250$ per arm.

When a smaller sample size of $n=100$ per arm is considered we still see excellent agreement between Rubin's variance estimator and the information anchored variance for up to 40–50% missing data (see Figure 3.7 with $\Delta = 0.3$; for $\Delta = 0$ and $\Delta = 1$ with $n=100$ per arm see Figures C.13 and C.14). This is also true for a larger sample size of $n=1000$ per arm (see Figure 3.8 with $\Delta = 0.3$; for $\Delta = 0$ and $\Delta = 1.0$ with $n=1000$ per arm see Figures C.15 and C.16). As expected, we see an even better approximation of the ideal variance by Rubin's variance estimator with a larger sample size.

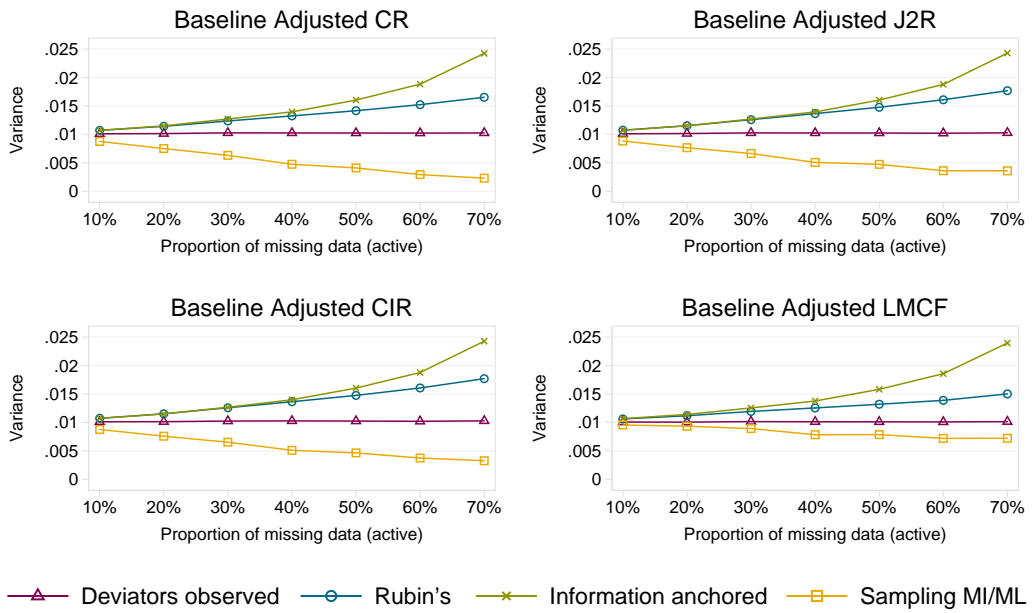


Figure 3.7: Rubin's baseline adjusted variance estimator vs. the baseline adjusted information anchored variance vs. baseline adjusted variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 baseline adjusted MI estimates, with the asthma trial treatment effect of $\Delta = 0.3$ and $n=100$ per arm.

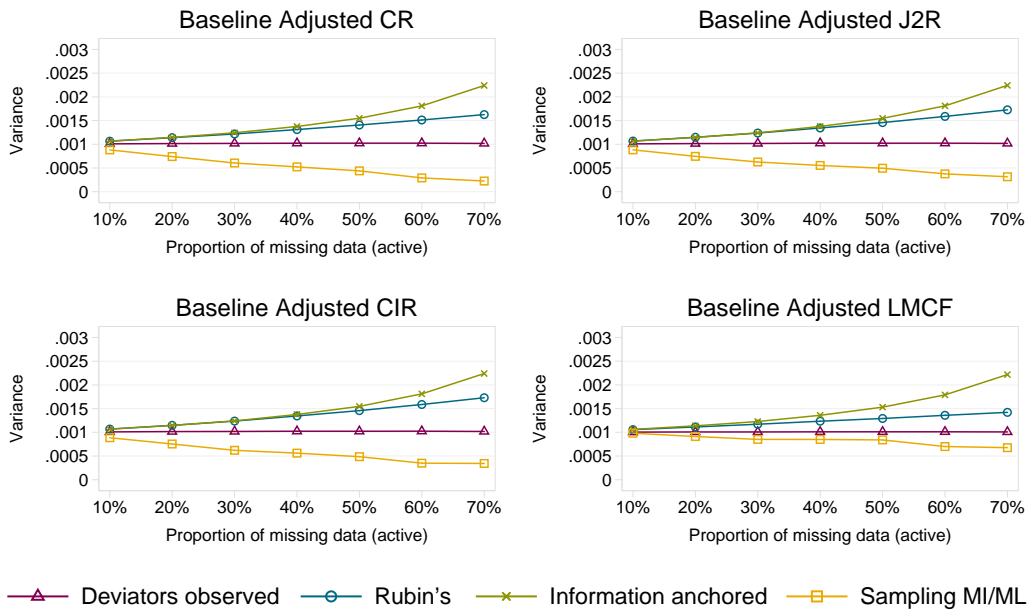


Figure 3.8: Rubin's baseline adjusted variance estimator vs. the baseline adjusted information anchored variance vs. baseline adjusted variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 baseline adjusted MI estimates, with the asthma trial treatment effect of $\Delta = 0.3$ and $n=1000$ per arm.

When we consider a low covariance between time 1 and time 2 measurements we see Rubin's variance still approximates the required variance well for up to 40–50% missingness in the active arm (see Figure 3.9 with $\Delta = 0.3$; for $\Delta = 0$ and $\Delta = 1$ with $\sigma_{12} = 0.05$ see Figures C.17 and C.18). When we consider a higher covariance between time 1 and time 2 measurements we see

Rubin's variance approximates the required variance well for even larger proportions of missing data (see Figure 3.10 with $\Delta = 0.3$; for $\Delta = 0$ and $\Delta = 1$ with $\sigma_{12} = 0.75$ see Figures C.19 and C.20). Agreement is excellent for up to 60–70% missingness.

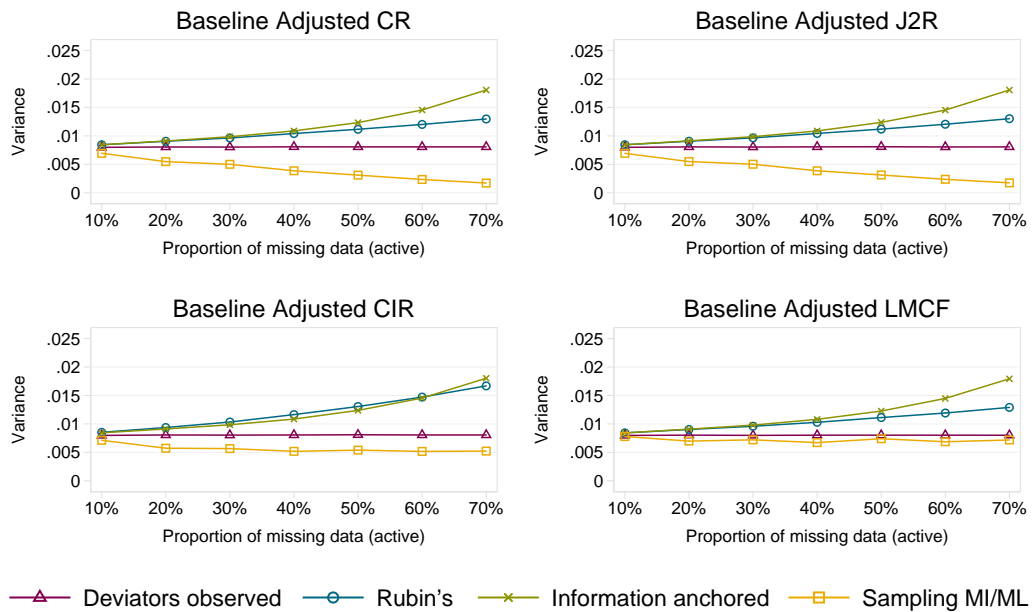


Figure 3.9: Rubin's baseline adjusted variance estimator vs. the baseline adjusted information anchored variance vs. baseline adjusted variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 baseline adjusted MI estimates, with the asthma trial treatment effect of $\Delta = 0.3$ and a low covariance between time 1 and time 2 outcomes ($\sigma_{12} = 0.05$). $n=250$ per arm.

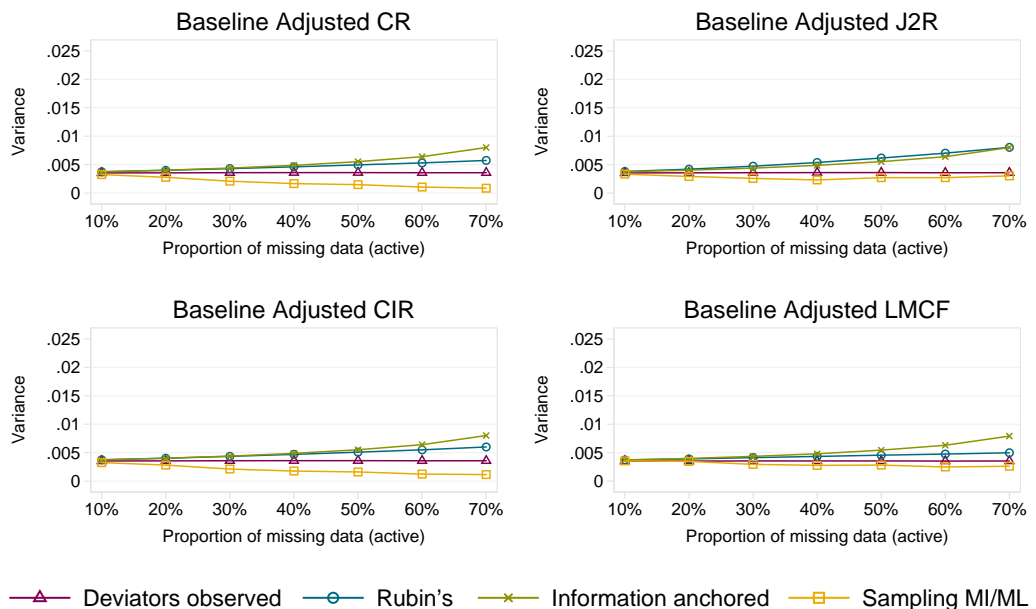


Figure 3.10: Rubin's baseline adjusted variance estimator vs. the baseline adjusted information anchored variance vs. baseline adjusted variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 baseline adjusted MI estimates, with the asthma trial treatment effect of $\Delta = 0.3$ and a high covariance between time 1 and time 2 outcomes ($\sigma_{12} = 0.75$). $n=250$ per arm.

Discussion

The agreement between Rubin's variance estimator and the information anchored variance for the baseline adjusted treatment effect is excellent in a variety of realistic trial settings (i.e. for up to 40-50% missing data in the active arm). The results confirm the findings from the derivations. Rubin's variance formula asymptotically approximates the information preserving variance. We conclude we are also justified in using Rubin's rules in the baseline adjusted single follow-up trial setting.

3.5 Summary

In this chapter we have defined our information anchoring principle for variance estimation in reference based sensitivity analysis. We have shown that Rubin's MI variance estimator approximates the information anchored variance we desire in the baseline and single follow-up trial setting for the unadjusted and baseline adjusted average treatment difference. That is, it approximates the design based variance estimate we would obtain, were we able to observe deviation data in the given scenario, inflated by the loss of information in the primary design based analysis. The difference is of a very small order, and negligible in realistic trial settings. For $n \rightarrow \infty$ it tends to zero faster than the actual information anchored variance. This provides justification for use of Rubin's variance formula in these settings. Rubin's variance estimator keeps the information loss constant across the primary and sensitivity analysis.

In this chapter our exploration has focused on the baseline and single follow-up trial setting with deviation in the active arm only. In the next chapter we present a general theory on the properties of Rubin's variance estimator in the baseline and single follow-up setting, then extend this for more complex longitudinal trial settings.

Chapter 4

General theory for Rubin's variance estimator in reference based analysis

So far we have introduced reference based sensitivity analysis, and argued that since the design based analysis model is retained in the sensitivity analysis, a natural principle for the treatment estimator variance is to keep the information loss due to missing data constant or *anchored* across the primary analysis and the sensitivity analyses. In this chapter we present a general theory of the information anchoring properties of Rubin's variance estimator for the baseline and single follow-up trial setting following reference based MI in Section 4.1. The substantive model considered here is a linear regression of a continuous outcome on randomised treatment group and observed baseline covariates. However, the general results presented can be applied to other settings which meet the outlined conditions.

We then extend the theory for more complex longitudinal trial settings. First for longitudinal trials where only the last measured variable is subject to non-response in Section 4.2. Then we consider longitudinal trials with more than one deviation pattern in Section 4.3. We restrict our attention to monotone missingness patterns since this is the most commonly observed pattern of missing data in longitudinal trials. A linear regression of a continuous outcome at the final time point on randomised treatment group and observed baseline covariates is the substantive analysis model of interest in the longitudinal settings.

The theory is initially developed under the assumption of an infinite number of imputations, K . In Section 4.4 we assess the impact of a finite K , as occurs in practice. In Section 4.5 we finish this chapter with a simulation study designed to explore the performance of Rubin's variance estimator in a variety of longitudinal trial settings.

As in Chapter 3 let $V_{\text{DJ, full}}$ denote the variance estimator for the treatment effect in the design based analysis with fully observed data if no deviations occur; $V_{\text{DF, full}}$ denotes the variance estimator for the treatment effect in the design based analysis where deviations occur off-treatment, but the deviation data is fully observed (under one of CR, J2R, CIR or LMCF); $V_{\text{DJ, MI}}$ denotes

Rubin’s variance estimator for the treatment effect in the primary design based analysis where post-deviation data are unobserved but imputed under our de-jure (on-treatment MAR) assumption and $V_{\text{DF, MI}}$ denotes Rubin’s variance estimator for the treatment effect in the design based sensitivity analysis where post-deviation data are unobserved but imputed under de-facto behaviour (one of CR, J2R, CIR or LMCF).

4.1 Baseline and single follow-up setting

Here we present a general theory on Rubin’s variance estimator and how it anchors the information loss due to missing data in the primary analysis across sensitivity analysis in the baseline and single follow-up reference based MI setting. For ease of exposition the initial proof focuses on the setting with deviation in one arm only. In Subsection 4.1.6 we show how the result holds with deviation in both arms. We also consider an infinite number of imputations, K . In Section 4.4 we show how the results in this chapter hold with finite K .

4.1.1 Setting for Proposition 1

Consider a two arm trial consisting of $2n$ patients with a single continuous follow-up outcome and baseline. n patients have been randomised to an active arm and n to a reference arm. \mathbf{Y} is the $(2n \times 1)$ vector of outcome data, and \mathbf{x} , is the corresponding $(2n \times p)$ matrix of observed baseline covariates including baseline outcome and treatment. The estimated treatment effect can be written $\mathbf{a}^T \mathbf{Y}$ where \mathbf{a} may or may not incorporate \mathbf{x} . For example, when the substantive model is a linear regression model $\mathbf{a}^T = (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}^T$.

We assume \mathbf{Y} is normally distributed and has a known variance Σ . The variance of the data is assumed to be known i.e. we do not incorporate a prior distribution on the variance in imputation without any serious loss of generality, following inference by Carpenter and Kenward in [44] (p. 56–59). Carpenter and Kenward examine Rubin’s rules for a simple mean with known and unknown variance and show, because of the conditional argument underlying the derivation, the additional complication of an unknown variance does not have a material effect on the development. It simply results in multiples of terms that are ≈ 1 . Thus, results are asymptotically the same.

Suppose all n reference patients and n_o of the active arm are observed without deviation at baseline and follow-up. The remaining n_d active arm patients deviate post-baseline such that for the active arm $n_o + n_d = n$. Let \mathcal{D} and \mathcal{O} define the sets of indices for the deviators and completers in the active arm respectively and assume the variance-covariance matrix of baseline and follow-up is the same in both treatment arms.

Assume $V_{\text{DF, full}} = V_{\text{DJ, full}} + O(n^{-2})$. Typically, although not necessarily, some form of regression model will be the substantive model of interest, used to make inferences about the treatment effect in the population of interest. As shown in Appendix B.1, for an unadjusted mean difference where $\pi_d = n_d/n$,

$$E[V_{\text{DF, full}}] = \frac{2\sigma_{22}}{n} + \frac{\Delta^2 \pi_d (1 - \pi_d)}{n},$$

where σ_{22} is the variance of the follow-up data, $\Delta = \mu_{a2} - \mu_{d2}$ and μ_{d2} represents the proposed mean at follow-up under the particular de-facto scenario of interest. For a baseline adjusted mean difference (as derived in Subsection 3.4.1),

$$E[V_{\text{DF, full}}] = \frac{2\sigma_{2.1}}{n} + \frac{\Delta^2 \pi_d (1 - \pi_d)}{n},$$

where $\sigma_{2.1}$ is the conditional variance of the follow-up data, given baseline. As discussed in Subsection 3.2.1, for a typical RCT powered at 80% with 5% statistical significance $\Delta^2 \approx 15.68\sigma_{22}/n$. Therefore $V_{\text{DF, full}} = V_{\text{DJ, full}} + O(n^{-2})$ holds for the mean treatment group difference, adjusted and unadjusted for baseline outcome i.e. holds for ANCOVA.

4.1.2 Proposition 1

For sensitivity analyses of a two arm trial with baseline and a single follow-up performed by MI using the algorithm of Carpenter, Roger and Kenward [13], the difference between Rubin's variance estimator and the information anchored variance for the treatment effect is,

$$\pi_d^2 \bar{\mathbf{P}}_{a,d} (\mathbf{V}_{a,o} - \mathbf{V}_{\text{DF},o}) \bar{\mathbf{P}}_{a,d}^T + \frac{E[\hat{B}_{\text{DJ}}]}{E[\hat{W}_{\text{DJ}}]} [O(n^{-2})].$$

$\mathbf{V}_{a,o}$ represents the variance-covariance matrix of the parameter estimates in the de-jure imputation model and $\mathbf{V}_{\text{DF},o}$ represents the variance-covariance matrix of the parameter estimates in the de-facto imputation model. $\bar{\mathbf{P}}_{a,d}$ is a vector containing the mean of the baseline covariates for each deviating active patient. In other words,

$$\frac{V_{\text{DJ, MI}}}{V_{\text{DJ, full}}} = \frac{V_{\text{DF, MI}}}{V_{\text{DF, full}}} + \pi_d^2 \bar{\mathbf{P}}_{a,d} (\mathbf{V}_{a,o} - \mathbf{V}_{\text{DF},o}) \bar{\mathbf{P}}_{a,d}^T + \frac{E[\hat{B}_{\text{DJ}}]}{E[\hat{W}_{\text{DJ}}]} [O(n^{-2})].$$

4.1.3 Proof of Proposition 1

Let $z = a, r$ index the randomised active arm or reference arm allocation for each patient i with follow-up outcome denoted by Y_{zi} . The outcome data for the reference patients is contained in the vector $\mathbf{Y}_r = (Y_{r1}, \dots, Y_{rn})^T$. The outcome data for the non-deviating active patients is contained in

the vector $\mathbf{Y}_{a,o} = \{Y_{ai}; i \in \mathcal{O}\}^T$.

We suppose that each deviating patient has two potential outcomes: the one that would occur if they remain on active treatment (de-jure) and the other that would occur off-treatment (de-facto). The potentially observable de-jure outcome data for the n_d deviating patients are contained in the vector $\mathbf{Y}_{a,DJ,d}$ and the alternative de-facto outcome data in the vector $\mathbf{Y}_{a,DF,d}$. Define $\mathbf{Y} = (\mathbf{Y}_r, \mathbf{Y}_{a,o}, \mathbf{Y}_{a,DJ,d}, \mathbf{Y}_{a,DF,d})^T$ as the collection of observed and potentially observable outcome data, which has dimensions $[(n + n_o + 2n_d) \times 1]$.

For each deviating patient we can only observe one of the potential outcomes, either de-jure or de-facto. Consider two $[(n + n_o + 2n_d) \times (n + n_o + 2n_d)]$ matrices, \mathbf{D}_{DJ} and \mathbf{D}_{DF} of 0's and 1's, arranged such that $\mathbf{D}_{DJ}\mathbf{Y}$ gives the $[(n + n_o + 2n_d) \times 1]$ de-jure data (with zero entries for the potential de-facto outcomes). $\mathbf{D}_{DF}\mathbf{Y}$ gives the $[(n + n_o + 2n_d) \times 1]$ de-facto data (with zero entries for the potential de-jure outcomes).

Let \mathbf{a} be a $[(n + n_o + 2n_d) \times 1]$ vector such that $\mathbf{a}^T\mathbf{D}_{DJ}\mathbf{Y}$ returns the de-jure treatment estimate and $\mathbf{a}^T\mathbf{D}_{DF}\mathbf{Y}$ returns the de-facto treatment estimate. When deviating patients experience on-treatment/de-jure behaviour post-deviation and their follow-up outcome data are observed the expectation of the variance estimator for the de-jure on-treatment estimand can be expressed as,

$$E[V_{DJ, \text{full}}] = E[V(\mathbf{a}^T\mathbf{D}_{DJ}\mathbf{Y})] = E[\mathbf{a}^T\mathbf{D}_{DJ}V(\mathbf{Y})\mathbf{D}_{DJ}^T\mathbf{a}] = \mathbf{a}^T\mathbf{D}_{DJ}\Sigma\mathbf{D}_{DJ}^T\mathbf{a}. \quad (4.1)$$

We use the proposition condition that the variance estimator for the de-facto estimand can be expressed as,

$$E[V_{DF, \text{full}}] = \mathbf{a}^T\mathbf{D}_{DJ}\Sigma\mathbf{D}_{DJ}^T\mathbf{a} + O(n^{-2}). \quad (4.2)$$

We now suppose that post-deviation data are unobserved, i.e. the potentially observable de-facto and de-jure entries in \mathbf{Y} are missing for the n_d active patients. We alternatively impute these outcomes, using de-jure imputation and de-facto imputation. This gives K 'complete' data samples \mathbf{Y}_k , of size $[(n + n_o + 2n_d) \times 1]$. For this we need an appropriate imputation model for the missing data under each scenario, with a suitable posterior for the included parameters.

Under our de-jure assumption (on-treatment MAR) our imputation model is formed from the regression of $\mathbf{Y}_{a,o}$ on $\mathbf{P}_{a,o}$ where $\mathbf{P}_{a,o}$ is the $[n_o \times p]$ design matrix for the imputation model, which contains the values of the $(p-1)$ covariates included in the imputation model (including the baseline outcome but excluding the treatment indicator since we perform imputation separately by arm) with a vector of 1's to include an intercept term in the model, for observed active patients.

The parameter estimates for the de-jure imputation model are found using, $\hat{\beta}_{a,o} = (\mathbf{P}_{a,o}^T\mathbf{P}_{a,o})^{-1}\mathbf{P}_{a,o}^T\mathbf{Y}_{a,o}$ with assumed known covariance matrix $\mathbf{V}_{a,o} = (\mathbf{P}_{a,o}^T\mathbf{P}_{a,o})^{-1}\sigma^2$. We assume the large sample posterior for the parameter estimates of the de-jure imputation model, denoted as $\hat{\beta}_{DJ}$, is normal and centered on the ML estimator $\hat{\beta}_{a,o}$ with covariance matrix $\mathbf{V}_{a,o}$.

That is,

$$\hat{\boldsymbol{\beta}}_{DJ} | \mathbf{Y}_{a,o} \sim N(\hat{\boldsymbol{\beta}}_{a,o}; \mathbf{V}_{a,o}).$$

The de-jure imputation model for patient i and imputation k can therefore be expressed as,

$$\check{Y}_{ai,k} | \mathbf{Y}_{a,o} = \mathbf{P}_{a,d,i} \left[\hat{\boldsymbol{\beta}}_{a,o} + \mathbf{b}_{a,o,k} \right] + e_{i,k} \text{ for } i \in \mathcal{D}, \mathbf{b}_{a,o,k} \sim N(0, \mathbf{V}_{a,o}), e_{i,k} \sim N(0, \sigma^2),$$

where $\mathbf{P}_{a,d,i}$ is the covariate data for each deviating active patient i (excluding treatment group but including a 1), of dimension $[1 \times p]$.

The conditional model for post-deviation data given pre-deviation data may take various different forms for de-facto imputation. Generally we assume the large sample posterior for the parameters of the imputation model, $\hat{\boldsymbol{\beta}}_{DF}$, is normal and centered on the ML estimator $\hat{\boldsymbol{\beta}}_{DF,o}$ with known covariance matrix $\mathbf{V}_{DF,o}$, that is,

$$\hat{\boldsymbol{\beta}}_{DF} | \mathbf{Y}_{DF,o} \sim N(\hat{\boldsymbol{\beta}}_{DF,o}; \mathbf{V}_{DF,o}),$$

where $\mathbf{Y}_{DF,o}$ consists of the relevant observed data under the particular de-facto setting of interest. The de-facto imputation model for patient i and imputation k can therefore be expressed as,

$$\check{Y}_{ai,k} | \mathbf{Y}_{DF,o} = \mathbf{P}_{a,d,i} \left[\hat{\boldsymbol{\beta}}_{DF,o} + \mathbf{b}_{DF,o,k} \right] + e_{i,k} \text{ for } i \in \mathcal{D}, \mathbf{b}_{DF,o,k} \sim N(0, \mathbf{V}_{DF,o}), e_{i,k} \sim N(0, \sigma^2).$$

Under the assumption of equal variance-covariance matrix of baseline and follow-up by treatment arm we consequently assume the same variance for the residuals in the de-jure and de-facto imputation models (denoted σ^2). In Subsection 4.1.5 we explore the impact of relaxing this assumption.

We are interested in imputation inference for, $\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_{DJ} \mathbf{Y}_k$ or $\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_{DF} \mathbf{Y}_k$.

Letting the number of imputations, $K \rightarrow \infty$, the variance of our MI treatment estimate as estimated by Rubin's rules is, $V_{DJ, MI} = \hat{W}_{DJ} + \hat{B}_{DJ}$ or $V_{DF, MI} = \hat{W}_{DF} + \hat{B}_{DF}$ where under the conditions required in the proposition,

$$E \left[\hat{W}_{DJ} \right] = E \left[\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_{DJ} \hat{\boldsymbol{\Sigma}}_k \mathbf{D}_{DJ}^T \mathbf{a} \right] \rightarrow \mathbf{a}^T \mathbf{D}_{DJ} \boldsymbol{\Sigma} \mathbf{D}_{DJ}^T \mathbf{a},$$

$$E \left[\hat{W}_{\text{DF}} \right] = E \left[\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_{\text{DF}} \hat{\Sigma}_k \mathbf{D}_{\text{DF}}^T \mathbf{a} \right] \rightarrow \mathbf{a}^T \mathbf{D}_{\text{DF}} \Sigma \mathbf{D}_{\text{DF}}^T \mathbf{a} + O(n^{-2}).$$

Under de-jure,

$$\hat{B}_{\text{DJ}} = \frac{1}{K-1} \sum_{k=1}^K \left[\pi_d (\bar{e}_k - \bar{e}) + \pi_d (\bar{\mathbf{P}}_{a,d} \mathbf{b}_{a,o,k} - \bar{\mathbf{P}}_{a,d} \bar{\mathbf{b}}_{a,o}) \right]^2,$$

where $\bar{e}_k = \frac{1}{n_d} \sum_{i \in \mathcal{D}} e_{i,k}$, $\bar{e} = \frac{1}{K} \sum_{k=1}^K \bar{e}_k$, $\bar{\mathbf{P}}_{a,d} = \frac{1}{n_d} \sum_{i \in \mathcal{D}} \mathbf{P}_{a,d,i}$ and $\bar{\mathbf{b}}_{a,o} = \frac{1}{K} \sum_{k=1}^K \mathbf{b}_{a,o,k}$. This has expectation,

$$\begin{aligned} E \left[\hat{B}_{\text{DJ}} \right] &= \frac{1}{K-1} \left(\pi_d^2 \left[(K-1) \frac{\sigma^2}{n_d} + (K-1) \bar{\mathbf{P}}_{a,d} \mathbf{V}_{a,o} \bar{\mathbf{P}}_{a,d}^T \right] \right) \\ &= \pi_d^2 \left[\frac{\sigma^2 + n_d \bar{\mathbf{P}}_{a,d} \mathbf{V}_{a,o} \bar{\mathbf{P}}_{a,d}^T}{n_d} \right]. \end{aligned}$$

Under de-facto,

$$\hat{B}_{\text{DF}} = \frac{1}{K-1} \sum_{k=1}^K \left[\pi_d (\bar{e}_k - \bar{e}) + \pi_d (\bar{\mathbf{P}}_{a,d} \mathbf{b}_{\text{DF},o,k} - \bar{\mathbf{P}}_{a,d} \bar{\mathbf{b}}_{\text{DF},o}) \right]^2,$$

where $\bar{\mathbf{b}}_{\text{DF},o} = \frac{1}{K} \sum_{k=1}^K \mathbf{b}_{\text{DF},o,k}$. This has expectation,

$$\begin{aligned} E \left[\hat{B}_{\text{DF}} \right] &= \frac{1}{K-1} \left(\pi_d^2 \left[(K-1) \frac{\sigma^2}{n_d} + (K-1) \bar{\mathbf{P}}_{a,d} \mathbf{V}_{\text{DF},o} \bar{\mathbf{P}}_{a,d}^T \right] \right) \\ &= \pi_d^2 \left[\frac{\sigma^2 + n_d \bar{\mathbf{P}}_{a,d} \mathbf{V}_{\text{DF},o} \bar{\mathbf{P}}_{a,d}^T}{n_d} \right]. \end{aligned}$$

The information anchored variance is,

$$E [V_{\text{anchored}}] = \frac{E [V_{\text{DF, full}}] \left(E \left[\hat{W}_{\text{DJ}} \right] + E \left[\hat{B}_{\text{DJ}} \right] \right)}{E [V_{\text{DJ, full}}]} = E [V_{\text{DF, full}}] \left[1 + \frac{E \left[\hat{B}_{\text{DJ}} \right]}{E \left[\hat{W}_{\text{DJ}} \right]} \right].$$

Since $E \left[\hat{W}_{\text{DJ}} \right] = E [V_{\text{DJ, full}}]$ that is,

$$\begin{aligned}
E[V_{\text{anchored}}] &= \mathbf{a}^T \mathbf{D}_{\text{DJ}} \boldsymbol{\Sigma} \mathbf{D}_{\text{DJ}}^T \mathbf{a} + O(n^{-2}) + \frac{E[\hat{B}_{\text{DJ}}]}{E[\hat{W}_{\text{DJ}}]} [\mathbf{a}^T \mathbf{D}_{\text{DJ}} \boldsymbol{\Sigma} \mathbf{D}_{\text{DJ}}^T \mathbf{a} + O(n^{-2})] \\
&= \mathbf{a}^T \mathbf{D}_{\text{DJ}} \boldsymbol{\Sigma} \mathbf{D}_{\text{DJ}}^T \mathbf{a} + O(n^{-2}) + E[\hat{B}_{\text{DJ}}] + \frac{E[\hat{B}_{\text{DJ}}]}{E[\hat{W}_{\text{DJ}}]} O(n^{-2}). \tag{4.3}
\end{aligned}$$

If Rubin's variance estimator preserves the information loss due to missing data in the primary analysis (under MAR) in the sensitivity analysis then,

$$E[\hat{W}_{\text{DF}}] + E[\hat{B}_{\text{DF}}] \approx \mathbf{a}^T \mathbf{D}_{\text{DJ}} \boldsymbol{\Sigma} \mathbf{D}_{\text{DJ}}^T \mathbf{a} + O(n^{-2}) + E[\hat{B}_{\text{DJ}}] + \frac{E[\hat{B}_{\text{DJ}}]}{E[\hat{W}_{\text{DJ}}]} O(n^{-2}).$$

That is,

$$\begin{aligned}
\mathbf{a}^T \mathbf{D}_{\text{DJ}} \boldsymbol{\Sigma} \mathbf{D}_{\text{DJ}}^T \mathbf{a} + O(n^{-2}) + E[\hat{B}_{\text{DF}}] &\approx \mathbf{a}^T \mathbf{D}_{\text{DJ}} \boldsymbol{\Sigma} \mathbf{D}_{\text{DJ}}^T \mathbf{a} + O(n^{-2}) \\
&\quad + E[\hat{B}_{\text{DJ}}] + \frac{E[\hat{B}_{\text{DJ}}]}{E[\hat{W}_{\text{DJ}}]} O(n^{-2}).
\end{aligned}$$

After simplification and rearrangement this becomes,

$$0 \approx E[\hat{B}_{\text{DJ}}] - E[\hat{B}_{\text{DF}}] + \frac{E[\hat{B}_{\text{DJ}}]}{E[\hat{W}_{\text{DJ}}]} [O(n^{-2})]. \tag{4.4}$$

Which is,

$$0 \approx \pi_d^2 \bar{\mathbf{P}}_{a,d} (\mathbf{V}_{a,o} - \mathbf{V}_{\text{DF},o}) \bar{\mathbf{P}}_{a,d}^T + \frac{E[\hat{B}_{\text{DJ}}]}{E[\hat{W}_{\text{DJ}}]} [O(n^{-2})]. \tag{4.5}$$

This gives the required result,

$$\frac{V_{\text{DJ, MI}}}{V_{\text{DJ, full}}} = \frac{V_{\text{DF, MI}}}{V_{\text{DF, full}}} + \pi_d^2 \bar{\mathbf{P}}_{a,d} (\mathbf{V}_{a,o} - \mathbf{V}_{\text{DF},o}) \bar{\mathbf{P}}_{a,d}^T + \frac{E[\hat{B}_{\text{DJ}}]}{E[\hat{W}_{\text{DJ}}]} [O(n^{-2})].$$

□

This result shows the difference between Rubin's variance estimator and the information anchoring variance is a small quantity, under the described conditions. In comparison to the information anchoring variance we desire (4.3) it is practically negligible.

Although not immediately obvious, in practice the difference between the variance of the coefficients in the de-jure and de-facto imputation distributions, $\mathbf{V}_{a,o} - \mathbf{V}_{\text{DF},o}$, will be small. $\mathbf{V}_{a,o}$ and $\mathbf{V}_{\text{DF},o}$ are two very similar quantities. We know $\mathbf{V}_{a,o} = \mathbf{A}/n_o$ where \mathbf{A} is a generic variance-covariance matrix of the data. $\mathbf{V}_{\text{DF},o}$ will depend on the specific de-facto scenario. But as detailed in Section 2.1, the parameters for the imputation models in each de-facto setting are pieced together using the parameters from the two treatment arm MVN MAR models. Thus the variance of the coefficients in the de-facto imputation distribution $\mathbf{V}_{\text{DF},o}$, will not be too dissimilar to the variance of the coefficients in the de-jure imputation distribution in the current setting. Further in all de-facto cases $\mathbf{V}_{a,o} - \mathbf{V}_{\text{DF},o}$ is multiplied by π_d^2 . For realistic proportions of missing data seen in practice ($\leq 50\%$), π_d^2 will be $\leq 1/4$. Thus the first term on the RHS of (4.5) will be relatively small in comparison to the terms which dominate the composition of the information anchored variance we are targeting (4.3). The information anchored variance is dominated by $E[V_{\text{DJ, full}}] = E[\hat{W}_{\text{DJ}}] = \mathbf{a}^T \mathbf{D}_{\text{DJ}} \boldsymbol{\Sigma} \mathbf{D}_{\text{DJ}}^T \mathbf{a}$, which is $O(n^{-1})$.

The second term on the RHS of (4.5) consists of a quantity which is $O(n^{-2})$ multiplied by $E[\hat{B}_{\text{DJ}}]/E[\hat{W}_{\text{DJ}}]$. When there is no missing data $\hat{B}_{\text{DJ}} = 0$. The proportion of information with missing data (assuming $K \rightarrow \infty$), relative to the information in full data, denoted by P , can be estimated as,

$$P = \left(\frac{\hat{W}_{\text{DJ}}}{\hat{W}_{\text{DJ}} + \hat{B}_{\text{DJ}}} \right).$$

This implies $\hat{B}_{\text{DJ}} = \hat{W}_{\text{DJ}}(1 - P)/P$ which indicates, $\hat{B}_{\text{DJ}} < \hat{W}_{\text{DJ}}$ for $P > 0.5$. So unless the information available with missing data is $\leq 1/2$ that in the full data, \hat{B}_{DJ} will be less than \hat{W}_{DJ} . In practice we would not expect to lose more than half of the intended information. Since $E[\hat{B}_{\text{DJ}}]$ is typically much smaller than $E[\hat{W}_{\text{DJ}}]$ for realistic trial scenarios i.e. with small to moderate proportions of missing data, the second term on the RHS will also be of a relatively small size.

To help understand this result we look at the CR example and then make some more general remarks.

Copy reference example

Suppose $Y_{zi} \stackrel{iid}{\sim} N(\mu_z, \sigma_{22})$ for $z = a, r$; the deviating patients CR and interest lies in the mean treatment group difference in Y . In this setting we know,

$$E[V_{\text{DJ, full}}] = \frac{2\sigma_{22}}{n}, E[V_{\text{CR, full}}] = \frac{2\sigma_{22}}{n} + \frac{\Delta^2 \pi_d (1 - \pi_d)}{n},$$

where $\Delta = \mu_a - \mu_r$ (see Appendix B.1 for derivation of this result). The CR imputation model is formed from the regression of \mathbf{Y}_r on \mathbf{P}_r where \mathbf{P}_r is the $[n \times p]$ design matrix for the imputation model, which contains the values of the $(p-1)$ covariates included in the imputation model (including the baseline outcome but excluding the treatment indicator) with a vector of 1's to include an intercept term in the model, for the n patients in the reference arm.

Parameter estimates for the CR de-facto imputation model are found as, $\hat{\boldsymbol{\beta}}_r = (\mathbf{P}_r^T \mathbf{P}_r)^{-1} \mathbf{P}_r^T \mathbf{Y}_r$ with assumed known covariance $\mathbf{V}_r = (\mathbf{P}_r^T \mathbf{P}_r)^{-1} \sigma^2$. We assume the large sample posterior for $\boldsymbol{\beta}_{\text{CR}}$ is normal and centered on the ML estimator $\hat{\boldsymbol{\beta}}_r$ with covariance matrix \mathbf{V}_r , that is,

$$\hat{\boldsymbol{\beta}}_{\text{CR}} | \mathbf{Y}_r \sim N(\hat{\boldsymbol{\beta}}_r; \mathbf{V}_r).$$

The CR imputation model for individual i for imputation k can therefore be expressed as,

$$\check{Y}_{ai,k} | \mathbf{Y}_r = \mathbf{P}_{a,d,i} [\hat{\boldsymbol{\beta}}_r + \mathbf{b}_{r,k}] + e_{i,k} \text{ for } i \in \mathcal{D} \text{ where, } \mathbf{b}_{r,k} \sim N(0, \mathbf{V}_r), e_{i,k} \sim N(0, \sigma^2).$$

For each imputation k we are interested in either,

$$\begin{aligned} \mathbf{a}^T \mathbf{D}_{\text{DJ}} \mathbf{Y}_k &= \frac{1}{n} \left[\sum_{i \in \mathcal{O}} Y_{ai} + \sum_{i \in \mathcal{D}} \check{Y}_{ai,k} - \sum_{i=1}^n Y_{ri} \right] \\ &= \frac{1}{n} \left[n_o \bar{Y}_{a,o} + n_d \bar{\mathbf{P}}_{a,d} \hat{\boldsymbol{\beta}}_{a,o} + n_d \bar{\mathbf{P}}_{a,d} \mathbf{b}_{a,o,k} + n_d \bar{e}_k - n \bar{Y}_r \right], \end{aligned}$$

or,

$$\begin{aligned} \mathbf{a}^T \mathbf{D}_{\text{DF}} \mathbf{Y}_k &= \frac{1}{n} \left[\sum_{i \in \mathcal{O}} Y_{ai} + \sum_{i \in \mathcal{D}} \check{Y}_{ai,k} - \sum_{i=1}^n Y_{ri} \right] \\ &= \frac{1}{n} \left[n_o \bar{Y}_{a,o} + n_d \bar{\mathbf{P}}_{a,d} \hat{\boldsymbol{\beta}}_r + n_d \bar{\mathbf{P}}_{a,d} \mathbf{b}_{r,k} + n_d \bar{e}_k - n \bar{Y}_r \right]. \end{aligned}$$

It follows that,

$$E \left[\hat{W}_{DJ} \right] = \frac{2\sigma_{22}}{n}, E \left[\hat{W}_{CR} \right] = \frac{2\sigma_{22}}{n} + \frac{\Delta^2 \pi_d (1 - \pi_d)}{n}.$$

Under de-jure deviation the between imputation variance is computed as,

$$\hat{B}_{DJ} = \frac{1}{K-1} \sum_{k=1}^K \left[\pi_d (\bar{e}_k - \bar{e}) + \pi_d (\bar{\mathbf{P}}_{a,d} \mathbf{b}_{a,o,k} - \bar{\mathbf{P}}_{a,d} \bar{\mathbf{b}}_{a,o}) \right]^2.$$

The expected value of this estimate is,

$$\begin{aligned} E \left[\hat{B}_{DJ} \right] &= \frac{1}{K-1} \left(\pi_d^2 \left[(K-1) \frac{\sigma^2}{n_d} + (K-1) \bar{\mathbf{P}}_{a,d} \mathbf{V}_{a,o} \bar{\mathbf{P}}_{a,d}^T \right] \right) \\ &= \pi_d^2 \left[\frac{\sigma^2 + n_d \bar{\mathbf{P}}_{a,d} \mathbf{V}_{a,o} \bar{\mathbf{P}}_{a,d}^T}{n_d} \right]. \end{aligned}$$

Under CR deviation the between imputation variance is computed as,

$$\hat{B}_{CR} = \frac{1}{K-1} \sum_{k=1}^K \left[\pi_d (\bar{e}_k - \bar{e}) + \pi_d (\bar{\mathbf{P}}_{a,d} \mathbf{b}_{r,k} - \bar{\mathbf{P}}_{a,d} \bar{\mathbf{b}}_r) \right]^2.$$

where $\bar{\mathbf{b}}_r = \frac{1}{K} \sum_{k=1}^K \mathbf{b}_{r,k}$. This has expected value,

$$\begin{aligned} E \left[\hat{B}_{CR} \right] &= \frac{1}{K-1} \left(\pi_d^2 \left[(K-1) \frac{\sigma^2}{n_d} + (K-1) \bar{\mathbf{P}}_{a,d} \mathbf{V}_r \bar{\mathbf{P}}_{a,d}^T \right] \right) \\ &= \pi_d^2 \left[\frac{\sigma^2 + n_d \bar{\mathbf{P}}_{a,d} \mathbf{V}_r \bar{\mathbf{P}}_{a,d}^T}{n_d} \right]. \end{aligned}$$

The information anchored variance we require under CR is found by substituting $E \left[\hat{B}_{DJ} \right]$, $E \left[\hat{W}_{DJ} \right]$ and $E \left[\hat{W}_{CR} \right]$ into (4.3),

$$\begin{aligned} E [V_{\text{anchored}}] &= \frac{2\sigma_{22}}{n} + \frac{\Delta^2 \pi_d (1 - \pi_d)}{n} + \pi_d^2 \left[\frac{\sigma^2}{n_d} + \bar{\mathbf{P}}_{a,d} \mathbf{V}_{a,o} \bar{\mathbf{P}}_{a,d}^T \right] \\ &\quad + \pi_d^2 \left[\frac{\sigma^2}{n_d} + \bar{\mathbf{P}}_{a,d} \mathbf{V}_{a,o} \bar{\mathbf{P}}_{a,d}^T \right] \frac{\Delta^2 \pi_d (1 - \pi_d)}{2\sigma_{22}}. \end{aligned} \quad (4.6)$$

To establish the information anchoring properties of Rubin's variance formula we substitute $E \left[\hat{B}_{DJ} \right]$, $E \left[\hat{B}_{CR} \right]$, $E \left[\hat{W}_{DJ} \right]$ and $E \left[\hat{W}_{CR} \right]$ into (4.4) which gives,

$$0 \approx \pi_d^2 \left[\bar{\mathbf{P}}_{a,d} (\mathbf{V}_{a,o} - \mathbf{V}_r) \bar{\mathbf{P}}_{a,d}^T \right] + \pi_d^2 \left[\frac{\sigma^2}{n_d} + \bar{\mathbf{P}}_{a,d} \mathbf{V}_{a,o} \bar{\mathbf{P}}_{a,d}^T \right] \frac{\Delta^2 \pi_d (1 - \pi_d)}{2\sigma_{22}}. \quad (4.7)$$

We separately consider the two terms on the RHS of (4.7) which make up the difference between Rubin's variance estimator and the information anchoring variance. We refer to the first term on the RHS of (4.7) as RHS1 and the second term as RHS2. The magnitude of RHS1 depends on the difference between the variance of the parameters in the CR de-facto imputation model and the variance of the parameters in the de-jure imputation model. In practice the difference between the variance of the coefficients in the imputation distributions will be small. Under the assumption of equal variance-covariance matrix by treatment arm we assume $\mathbf{V}_{a,o}$ is \mathbf{A}/n_o and \mathbf{V}_r is \mathbf{A}/n where \mathbf{A} is a generic variance-covariance matrix of the data. Thus where $\pi_d = n_d/n$,

$$\mathbf{V}_{a,o} - \mathbf{V}_r = \mathbf{A} \left(\frac{1}{n_o} - \frac{1}{n} \right) = \mathbf{A} \left(\frac{\pi_d}{n(1 - \pi_d)} \right) = \mathbf{A} \frac{1}{n} [\pi_d + \pi_d^2 + \pi_d^3 \dots].$$

This reveals $\mathbf{V}_{a,o} - \mathbf{V}_r$ is $O(n^{-1})$. RHS1 consists of $\mathbf{V}_{a,o} - \mathbf{V}_r$ multiplied by π_d^2 , so will be of an even smaller magnitude. $\pi_d < 1$ and for realistic proportions of missing data i.e. $\pi_d \leq 1/2$, $\pi_d^2 \leq 1/4$. The largest term in the desired information anchored variance (4.6) is $E \left[\hat{W}_{DJ} \right] = 2\sigma_{22}/n$. This is $O(n^{-1})$. Thus RHS1 will always be relatively small in comparison to the quantity we are seeking.

RHS2 can be broken down into two components. The first part of RHS2,

$$\pi_d^2 \frac{\sigma^2}{n_d} \frac{\Delta^2 \pi_d (1 - \pi_d)}{2\sigma_{22}} = \pi_d^2 \frac{\sigma^2 \Delta^2 (1 - \pi_d)}{2\sigma_{22} n},$$

is $O(n^{-2})$ for a typical RCT powered at 80% with 5% statistical significance. As discussed in Subsection 3.2.1, in such cases $\Delta^2 \approx (\mu_1 - \mu_2)^2 = 15.68\sigma_{22}/n$. The second part of RHS2,

$$\pi_d^2 \bar{\mathbf{P}}_{a,d} \mathbf{V}_{a,o} \bar{\mathbf{P}}_{a,d}^T \frac{\Delta^2 \pi_d (1 - \pi_d)}{2\sigma_{22}}$$

is also $O(n^{-2})$ for a typical RCT. Thus RHS2 will also be of a relatively small size in comparison to the information anchored variance we are seeking, which is dominated by $2\sigma_{22}/n$.

To explicitly see this we consider the most basic two arm trial scenario with no baseline and just a single follow-up. Here the data for active deviators are imputed using the mean outcome from the randomised active arm under de-jure behaviour. Or the mean outcome from the reference arm under CR de-facto behaviour. Thus $\bar{\mathbf{P}}_{a,d,i}$ is a single constant of 1, therefore $\bar{\mathbf{P}}_{a,d}$ is also a single

constant of 1. Figure 4.1 plots the information anchored variance against the two components that form the difference between Rubin’s MI variance estimator and the information anchored variance (RHS1 and RHS2) for various sized trials. In each setting we assume the RCT is powered at 80% with 5% statistical significance, i.e. $\Delta^2 \approx 15.68\sigma_{22}/n$. Each component is expressed generally as a proportion of the variance of the outcome data. In all scenarios we can clearly see the relative small size of RHS1 and RHS2.

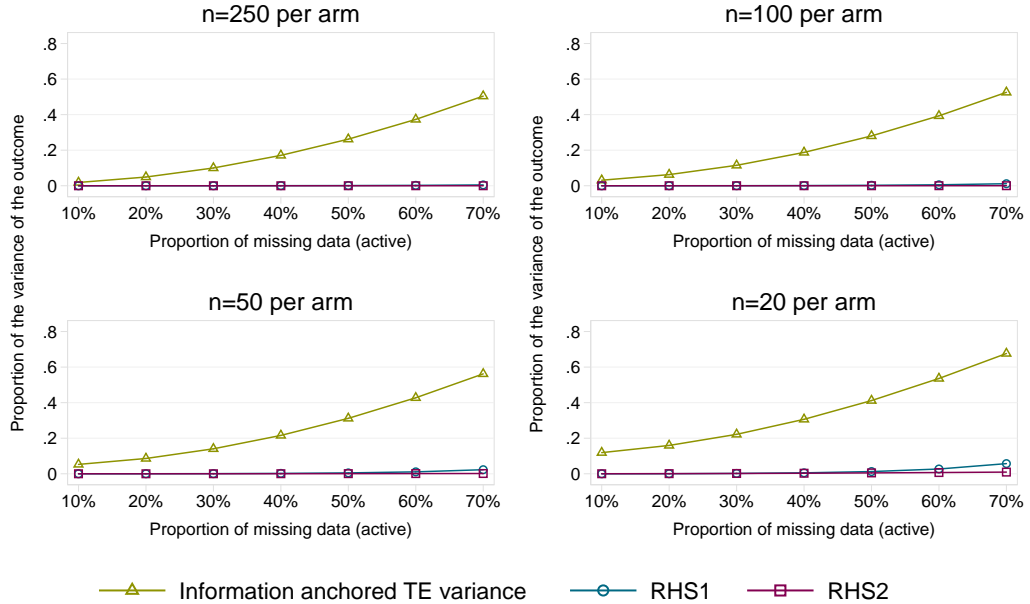


Figure 4.1: Information anchored variance vs. the difference between Rubin’s variance estimator and the information anchored variance (RHS1 and RHS2) for a trial with a single follow-up. Components are expressed as a proportion of the total variance of the data.

Proposition 1 corresponds with the inference in Subsection 3.2.1 where we explicitly derived all the components in this equation. In comparison to the information anchored variance we desire, the difference between Rubin’s variance estimator and the information anchored variance is in practice negligible. The results of the simulation study in Subsection 3.3.1 further demonstrate the small size of the components on the RHS of (4.7).

4.1.4 Implementation for improved information anchoring

We note that if $\mathbf{V}_{a,o} \approx \mathbf{V}_{DF,o}$ i.e. the variance of the parameters in the de-facto imputation model matches the variance of the parameters in the de-jure (MAR) imputation model, the approximation of the information anchored variance by Rubin’s variance estimator is sharpened. When this is the case the information anchoring (4.5) becomes,

$$0 \approx \frac{E \left[\hat{B}_{DJ} \right]}{E \left[\hat{W}_{DJ} \right]} \left[O(n^{-2}) \right]. \quad (4.8)$$

In the baseline and single follow-up setting under CR $\mathbf{V}_{a,o} \approx \mathbf{V}_r$ is obtained when the number of fully observed patients in the reference arm is close to n_o —the number of observed patients in the active arm at time 2— with the assumption of equal variance by arm.

If the number of observed active patients (n_o) and reference arm patients (n) are not similar in a trial with a single follow-up, to achieve improved information anchoring in practice, we can simply scale the variance of the parameters in the CR imputation distribution by n/n_o . With equal variance-covariance of the data by arm, which we denote by \mathbf{A} , $\mathbf{V}_{a,o} = \mathbf{A}/n_o$ and $\mathbf{V}_r = \mathbf{A}/n$. After scaling \mathbf{V}_r by n/n_o the variance of the parameters in the CR imputation model will therefore match the variance of the parameters in the de-jure (MAR) imputation model.

Alternatively prior to performing reference based imputation we can bootstrap the observed reference data in order to obtain a sample of the required size from the reference distribution, which will result in $\mathbf{V}_{a,o} \approx \mathbf{V}_r$. Bootstrapping involves repeated random sampling, with replacement, of observed data and enables the estimation of the sampling distribution of statistics of interest [22]. In the current context, repeated sampling of the observed reference data can be employed to estimate the distribution of the data in the reference arm. For a simple trial with single follow-up we can then draw a sample of n_o reference patients from the estimated reference distribution. The new sample of reference patients can then be used to construct the reference based imputation model, rather than the observed reference group patients.

For a trial with baseline (completely observed) and a single follow-up as explored in Proposition 1 after estimating the reference distribution via bootstrapping, we can draw a sample of n reference patients observed at baseline, of which n_o are also observed at time 2 to match the composition of the active group sample used in MAR imputation. The newly drawn sample can then be used to construct the reference based imputation model. This will also result in improved information preservation for the J2R and CIR procedures.

The variance of the parameters in LMCF imputation models will be closer to the variance of the parameters in de-jure (MAR) imputation models when the number of deviators is small, i.e. n_o is close to n . In the previous chapter we have already seen how the information anchoring property of Rubin’s rules is more accurate under LMCF imputation when the number of deviators is smaller.

Simulation study for improved information anchoring

To demonstrate the implementation for improved information anchoring in the CR, J2R and CIR settings we present results from an extension of the simulation study in Section 3.3.

Methods

The methods used to generate the datasets are described in Subsection 3.3.1. Interest lies in the mean treatment group difference at time 2. In this extension we consider the setting where $\Delta = 0.3$, as observed in the asthma trial, with 10–70% missing data at time 2 in the active arm only. As detailed in Subsection 3.3.1, for each missingness setting, the treatment effect and its estimated variance were computed using the full data. The analysis model was a linear regression of the time 2 response on randomised treatment group. In each scenario, following the imposition of the missing data, post-deviation data were regenerated using the appropriate conditional normal

distributions to compute the treatment effect and its variance had we observed the post-deviation data in the particular de-facto scenario.

We then altered the implementation of the reference based MI approach of Carpenter, Roger and Kenward [13]. First we drew an independent sample of n_o reference patients with baseline and time 2 data, and a further n_d with baseline data only from the known reference population to match the profile of the observed cases in the active arm. This sample was then used to build the required imputation models for CR, J2R and CIR imputation following the algorithm in [13], as outlined in Section 2.1. Of course, this approach is only possible in this simulation study set-up where the reference population distribution is known. Following imputation the sample of reference patients drawn for imputation were discarded from analysis. The analysis model was fitted to each imputed dataset in turn and results summarised for inference in the usual way using Rubin’s rules.

Secondly we bootstrapped the observed reference group data, using 1000 simple random draws with replacement, to estimate the mean and variance of the data in the reference group. We then drew an independent sample of n_o reference patients with baseline and week 12 data, and a further n_d with baseline data only from the estimated reference distribution. The bootstrap attempts to mimic, ‘going back to the population,’ which is not possible in practice. This second sample was then used to build the required imputation models for CR, J2R and CIR imputation following the algorithm in [13]. Following imputation the sample of reference patients drawn for imputation was again discarded from analysis. The analysis model was fitted to each imputed dataset in turn and results summarised for inference using Rubin’s rules.

The main outcomes of interest were Rubin’s MI variance estimate for the treatment effect under the alternative implementations, and the ideal information anchored variance estimate, computed as in Section 3.1. Estimates were averaged over the 1000 simulations in each scenario. The main focus was to show how a superior approximation of the information anchored variance can be obtained by Rubin’s MI variance estimator under the alternative implementations, in comparison to the standard reference based MI implementation (for which results were presented in Section 3.3.2).

Results

Figure 4.2 shows excellent information anchoring by Rubin’s variance estimator under CR for all missingness scenarios when an independent sample of reference group patients, chosen to match the profile of the observed active patients, is drawn from the known reference population to build the imputation models. This is as expected since $\mathbf{V}_{a,o} \approx \mathbf{V}_{DF,o}$. The approximation of the information anchored variance by Rubin’s variance corresponds with that seen for the standard CR MI implementation for up to 40–50% missing data in the active arm, but is notably improved for 50% or greater missingness. This is only possible via simulation.

However, Figure 4.2 also shows excellent information anchoring by Rubin’s variance estimator under CR for all missingness scenarios when an independent sample of reference group patients that matches the profile of the observed active patients is drawn via bootstrapping the observed data. This could be done in practice if desired.

Figures 4.3 and 4.4 show this also holds for J2R and CIR imputation. The improvement in the approximation for missingness amounts above 50% is not quite as perfect as under CR, but in both cases it is superior to the standard reference based MI implementation for larger amounts of

missingness.

Our focus is not on the long-run empirical sampling variance (over the 1000 simulations) of the MI estimator here. However it is interesting to see what affect these procedures have on it. We see that when we draw a sample of independent reference group patients from the population for imputation the behaviour of the long-run empirical sampling variance changes (Figures 4.2–4.4). As discussed in Subsection 2.2.2 when we conduct reference based MI using the observed reference group data we introduce a covariance between the mean of the deviating active patients and the observed reference group patients. A consequence of borrowing information between trial arms. When analysing the treatment group difference this covariance is taken away in the computation of the sampling variance. Because the sample of reference group patients used here in imputation is an independent draw from the known population, and not included in the analysis, the unwanted covariance is no longer incorporated.

However, the data from the reference patients that are drawn to build the imputation model are not included in the analysis. There is therefore a mismatch in the conditioning used in the analysis and the imputation. This explains why the long-run empirical sampling variance still does not match Rubin’s variance estimator, i.e. the information anchored variance we are targeting. This property is discussed by Reiter in [59].

In practice we would have to use the observed data to estimate the parameters of the reference distribution. When we bootstrap the observed data to estimate the reference distribution for imputation, the empirical long-run sampling variance of the MI estimate is also unsuitable. It is smaller than the variance estimate we would obtain if the off-treatment behaviour of the deviators was actually observed. Moreover, since the data used to build the imputation model are not included in the analysis (we analyse only the original randomised patients data) there is also a mismatch in the conditioning used in the analysis and the imputation [59]. This is therefore why we also see that the empirical long-run sampling variance does not correspond with Rubin’s variance estimate and the information anchored variance via the bootstrapping implementation.

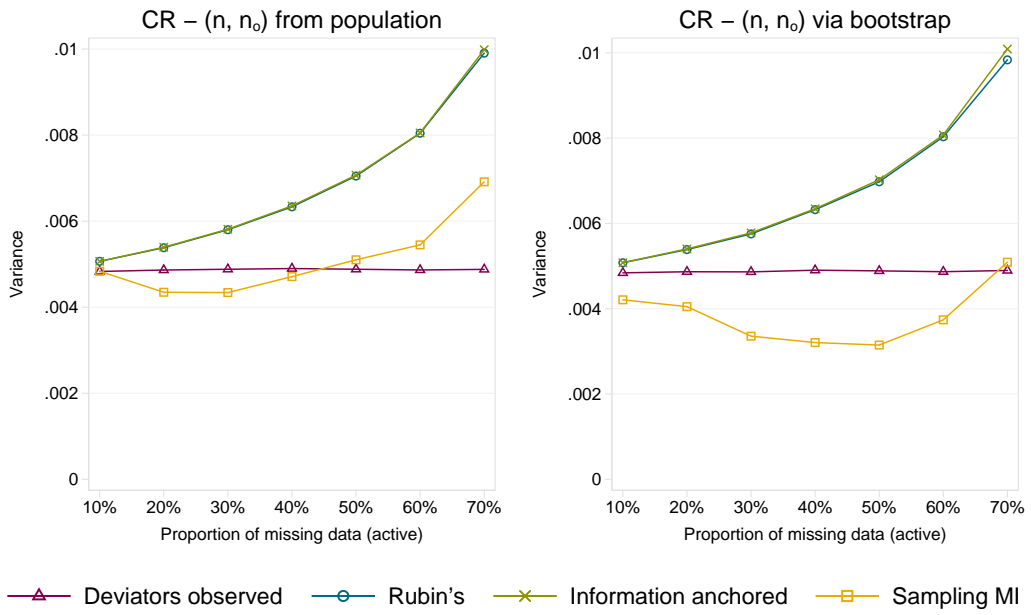


Figure 4.2: CR MI with implementation for improved information anchoring. Rubin's variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates. $\Delta = 0.3$ and $n=250$ per arm.

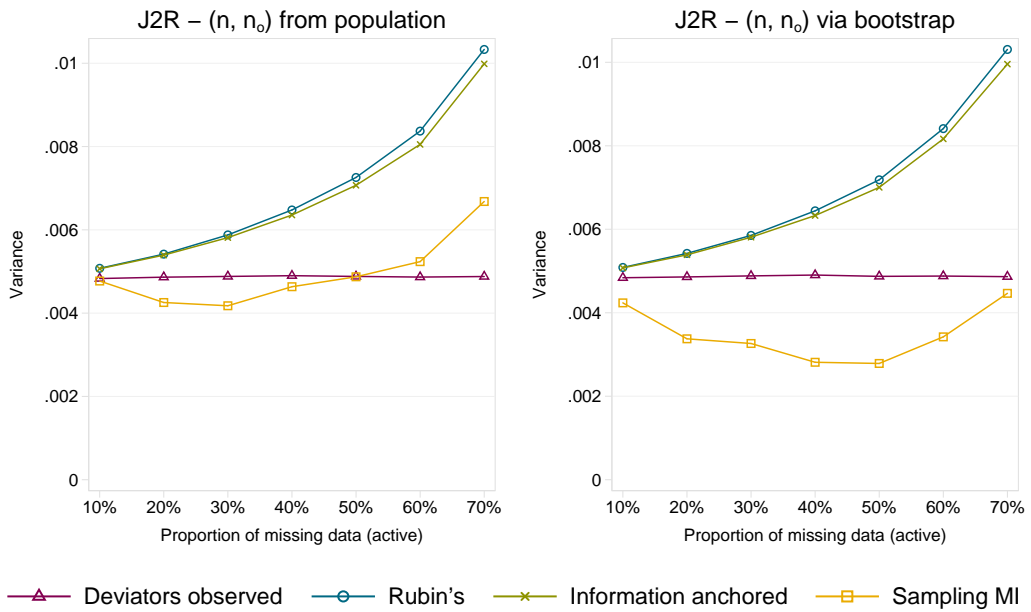


Figure 4.3: J2R MI with implementation for improved information anchoring. Rubin's variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates. $\Delta = 0.3$ and $n=250$ per arm.

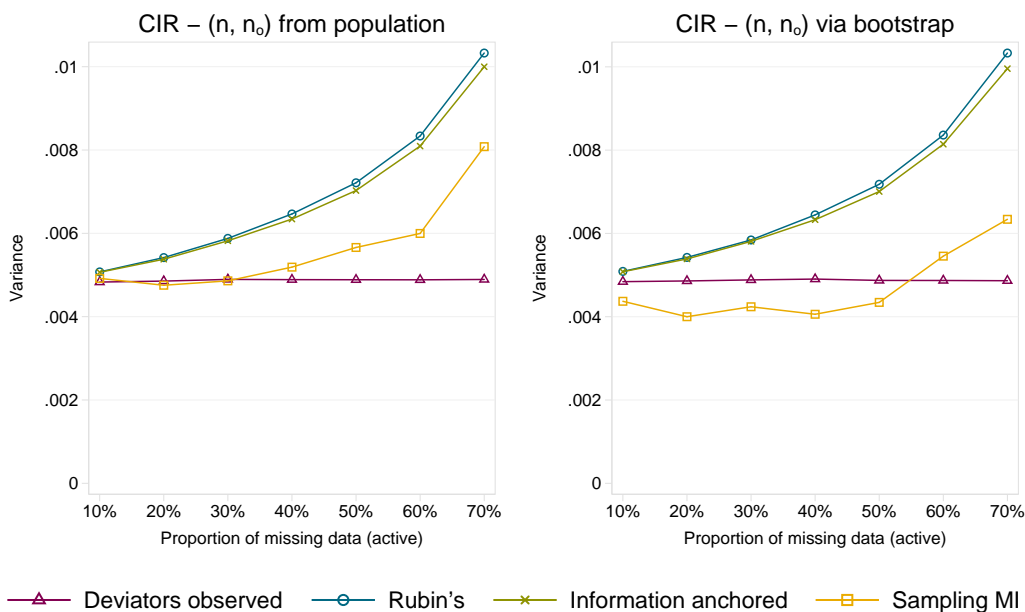


Figure 4.4: CIR MI with implementation for improved information anchoring. Rubin’s variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates. $\Delta = 0.3$ and $n=250$ per arm.

Discussion

Improved information anchoring can be achieved by Rubin’s variance estimator if reference based imputation is conducted using a sample of reference patients that has the same profile of the observed active arm patients. This is obtained via bootstrapping the observed reference data. However, little is gained over using the standard implementation for up to 40-50% missing data in the active arm, since as discussed in the previous chapter, Rubin’s variance estimator provides excellent information anchoring in these circumstances. A similar effect could be obtained by scaling the variance of the parameters in the de-facto imputation model.

Information anchoring will naturally be stronger for a trial with imputed reference based deviation data in an active arm when the number of observed patients in the active arm corresponds with the total number of reference patients. When the number of reference cases is considerably different to the number of observed active cases the proposed alternative implementation may be useful. However this is unlikely in practice.

4.1.5 Relaxing the equal variance assumption

Proposition 1 included the condition for equal variance-covariance matrix for baseline and follow-up between the trial arms. This is a common assumption made in the clinical trial setting. Here we consider the impact of unequal variance structures across the trial arms.

We follow the proof of Proposition 1 for settings in which $V_{DF, \text{full}} = V_{DJ, \text{full}} + O(n^{-2})$. If the equal variance assumption does not hold then the variance of the residuals ($e_{i,k}$) in the de-jure/on-

treatment MAR and de-facto/off-treatment imputation models, denoted by σ^2 , cannot be assumed to be the same. We therefore alternatively now denote the variance of the residuals in the de-jure imputation model as σ_{DJ}^2 and in the de-facto imputation model as σ_{DF}^2 . Then under de-jure imputation,

$$E \left[\hat{B}_{DJ} \right] = \pi_d^2 \left[\frac{\sigma_{DJ}^2 + n_d \bar{\mathbf{P}}_{a,d} \mathbf{V}_{a,o} \bar{\mathbf{P}}_{a,d}^T}{n_d} \right],$$

and under de-facto,

$$E \left[\hat{B}_{DF} \right] = \pi_d^2 \left[\frac{\sigma_{DF}^2 + n_d \bar{\mathbf{P}}_{a,d} \mathbf{V}_{DF,o} \bar{\mathbf{P}}_{a,d}^T}{n_d} \right].$$

The requirement for information anchoring becomes,

$$0 \approx \pi_d^2 \left[\frac{\sigma_{DJ}^2 - \sigma_{DF}^2}{n_d} + \bar{\mathbf{P}}_{a,d} (\mathbf{V}_{a,o} - \mathbf{V}_{DF,o}) \bar{\mathbf{P}}_{a,d}^T \right] + \frac{E \left[\hat{B}_{DJ} \right]}{E \left[\hat{W}_{DJ} \right]} [O(n^{-2})]. \quad (4.9)$$

We see that when we relax the equal variance assumption, the difference between Rubin's variance estimator and the information anchored variance remains small in size in comparison to the information anchored variance we are seeking. The information anchored variance is dominated by $E \left[\hat{W}_{DJ} \right]$, which is $O(n^{-1})$. There is now an additional component in the difference between Rubin's variance and the ideal variance, which is driven by the degree of difference in the variance structure by trial arm. The larger the difference in the variance structure by trial arm, the greater the difference between σ_{DF}^2 and σ_{DJ}^2 . However, in practice the variance structures are not likely to differ too drastically by arm (although there will always be exceptions). Further this additional component is always multiplied by π_d/n i.e. is $O(\pi_d n^{-1})$, thus will always be relatively small.

The difference between $\mathbf{V}_{a,o} - \mathbf{V}_{DF,o}$ will also be larger when the the variance structures differ by trial arm. However if we denote the variance-covariance matrix of the data in the active arm as \mathbf{A} , the variance-covariance matrix of the data in the reference arm can always be expressed as $\mathbf{A} + \mathbf{B}$ where \mathbf{B} is a matrix of constants. The additional difference can therefore be expressed as a component with the same order as $\mathbf{V}_{DF,o}$ in the equal variance setting. Which we would expect to be small anyway since \mathbf{A} and \mathbf{B} are not likely to be too different. Thus the extra difference does not effect the order of this component in (4.9). The term remains multiplied by π_d^2 . The inferences for the limiting behaviour of this term when the variance-covariance structures were equal still apply here.

Relaxing the equal variance assumption does not have a profound effect on the approximation between Rubin's variance estimator and the information anchored variance.

4.1.6 Extension for deviation in both arms

Up to here, the proof of Proposition 1 has focused on the two arm trial setting with deviation in one arm only. This is not a requirement for the approximation to hold, rather just a simplification for clarity. Below we explore the impact of deviation in both arms. We remain focused on settings where $V_{\text{DF, full}} = V_{\text{DJ, full}} + O(n^{-2})$. If $V_{\text{DF, full}} = V_{\text{DJ, full}} + O(n^{-2})$ with deviation in one arm then with deviation in both arms we will also obtain $V_{\text{DF, full}} = V_{\text{DJ, full}} + O(n^{-2})$, Appendix B.1 illustrates this for the baseline and single follow-up setting where the treatment effect is the mean difference in the follow-up outcome. Since an unequal variance structure by arm has a negligible effect in practice, we maintain the assumption of equal variance-covariance structure for baseline and follow-up in what follows.

Suppose among the n reference patients, only $n_{r,o}$ are actually observed, while the remaining $n_{r,d}$ deviate post-baseline. Let \mathcal{RD} and \mathcal{RO} define the sets of indices for the deviators and completers in the reference arm respectively. Let $\pi_{r,d} = n_{r,d}/n$. The outcome data for the observed reference patients are contained in the vector $\mathbf{Y}_{r,o} = \{Y_{ri}; i \in \mathcal{RO}\}^T$. The potentially observable de-jure data for the $n_{r,d}$ deviating reference patients are contained in the vector $\mathbf{Y}_{r,DJ,d}$ and the alternative de-facto outcome data in the vector $\mathbf{Y}_{r,DF,d}$. The full collection of observed and potentially observable outcome data is now defined as $\mathbf{Y} = (\mathbf{Y}_{r,o}, \mathbf{Y}_{r,DJ,d}, \mathbf{Y}_{r,DF,d}, \mathbf{Y}_{a,o}, \mathbf{Y}_{a,DJ,d}, \mathbf{Y}_{a,DF,d})^T$ which has dimensions $[(n_{r,o} + 2n_{r,d} + n_o + 2n_d)]$. We assume \mathbf{Y} is normally distributed and has known variance Σ .

We redefine the two matrices \mathbf{D}_{DJ} and \mathbf{D}_{DF} so that $\mathbf{D}_{\text{DJ}}\mathbf{Y}$ and $\mathbf{D}_{\text{DF}}\mathbf{Y}$ now each give the de-jure data or de-facto data across both treatment arms. We focus on settings where $E[V_{\text{DF, full}}] = \mathbf{a}^T \mathbf{D}_{\text{DJ}} \Sigma \mathbf{D}_{\text{DJ}}^T \mathbf{a} + O(n^{-2})$.

We follow the steps outlined in Section 4.1 to establish the de-jure imputation model for the deviating active arm patients. Under de-jure imputation, the imputation model for patient i and imputation k in the active arm is expressed as,

$$\tilde{Y}_{ai,k} | \mathbf{Y}_{a,o} = \mathbf{P}_{a,d,i} \left[\hat{\beta}_{a,o} + \mathbf{b}_{a,o,k} \right] + e_{i,k} \text{ for } i \in \mathcal{D},$$

where $\mathbf{b}_{a,o,k} \sim N(0, \mathbf{V}_{a,o})$, $e_{i,k} \sim N(0, \sigma^2)$ and $\mathbf{P}_{a,d,i}$ is the covariate data for each deviating active patient i (excluding treatment group but including a 1), of dimensions $[1 \times p]$.

For the reference arm, under our de-jure assumption (on-treatment MAR), our imputation model is formed from the regression of $\mathbf{Y}_{r,o}$ on $\mathbf{P}_{r,o}$ where $\mathbf{P}_{r,o}$ is the $[n_{r,o} \times p]$ design matrix for the imputation model, which contains the values of the $(p-1)$ covariates included in the imputation model (including the baseline outcome but excluding the treatment indicator since we perform imputation separately by arm) with a vector of 1's to include an intercept term in the model, for observed reference patients.

The parameter estimates for the de-jure imputation model are found using, $\hat{\beta}_{r,o} = (\mathbf{P}_{r,o}^T \mathbf{P}_{r,o})^{-1} \mathbf{P}_{r,o}^T \mathbf{Y}_{r,o}$ with assumed known covariance matrix $\mathbf{V}_{r,o} = (\mathbf{P}_{r,o}^T \mathbf{P}_{r,o})^{-1} \sigma^2$. We assume the large sample posterior for the parameter estimates for the de-jure imputation model,

denoted as $\hat{\beta}_{DJ,r}$, is normal and centered on the ML estimator $\hat{\beta}_{r,o}$ with covariance matrix $\mathbf{V}_{r,o}$, that is,

$$\hat{\beta}_{DJ,r} | \mathbf{Y}_{r,o} \sim N(\hat{\beta}_{r,o}; \mathbf{V}_{r,o}).$$

The de-jure imputation model for patient i and imputation k in the reference arm can therefore be expressed as,

$$\tilde{Y}_{ri,k} | \mathbf{Y}_{r,o} = \mathbf{P}_{r,d,i} \left[\hat{\beta}_{r,o} + \mathbf{b}_{r,o,k} \right] + e_{i,k} \text{ for } i \in \mathcal{RD},$$

where $\mathbf{b}_{r,o,k} \sim N(0, \mathbf{V}_{r,o})$, $e_{i,k} \sim N(0, \sigma^2)$ and $\mathbf{P}_{r,d,i}$ is the covariate data for each deviating reference patient i (excluding treatment group but including a 1), of dimensions $[1 \times p]$.

Under de-facto imputation for patients in the active arm we assume the large sample posterior for the parameters of the imputation model, which we denote by $\hat{\beta}_{DF,a}$, is normal and centered on the ML estimator $\hat{\beta}_{DF,a,o}$ with known covariance matrix $\mathbf{V}_{DF,a,o}$ that is,

$$\hat{\beta}_{DF,a} | \mathbf{Y}_{DF,a,o} \sim N \left(\hat{\beta}_{DF,a,o}, \mathbf{V}_{DF,a,o} \right),$$

where $\mathbf{Y}_{DF,a,o}$ consists of the relevant observed outcome data under the particular de-facto setting. The de-facto imputation model for patient i and imputation k in the active arm can therefore be expressed as,

$$\tilde{Y}_{ai,k} | \mathbf{Y}_{DF,a,o} = \mathbf{P}_{a,d,i} \left[\hat{\beta}_{DF,a,o} + \mathbf{b}_{DF,a,o,k} \right] + e_{i,k} \text{ for } i \in \mathcal{D},$$

where $\mathbf{b}_{DF,a,o,k} \sim N(0, \mathbf{V}_{DF,a,o})$, and $e_{i,k} \sim N(0, \sigma^2)$. Under de-facto imputation for patients in the reference arm we assume the large sample posterior for the parameters of the imputation model, which we denote by $\hat{\beta}_{DF,r}$, is normal and centered on the ML estimator $\hat{\beta}_{DF,r,o}$ with known covariance matrix $\mathbf{V}_{DF,r,o}$ that is,

$$\hat{\beta}_{DF,r} | \mathbf{Y}_{DF,r,o} \sim N \left(\hat{\beta}_{DF,r,o}, \mathbf{V}_{DF,r,o} \right),$$

where $\mathbf{Y}_{DF,r,o}$ consists of the relevant observed outcome data under the particular de-facto setting. The de-facto imputation model for patient i and imputation k in the reference arm can therefore be expressed as,

$$\tilde{Y}_{ri,k} | \mathbf{Y}_{DF,r,o} = \mathbf{P}_{r,d,i} \left[\hat{\beta}_{DF,r,o} + \mathbf{b}_{DF,r,o,k} \right] + e_{i,k} \text{ for } i \in \mathcal{RD},$$

where $\mathbf{b}_{DF,r,o,k} \sim N(0, \mathbf{V}_{DF,r,o})$, and $e_{i,k} \sim N(0, \sigma^2)$. We are interested in imputation inference for, $\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_{DJ} \mathbf{Y}_k$ or $\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_{DF} \mathbf{Y}_k$. For Rubin's variance estimator, under the described conditions, $E \left[\hat{W}_{DJ} \right] = E \left[\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_{DJ} \hat{\Sigma}_k \mathbf{D}_{DJ}^T \mathbf{a} \right] \rightarrow \mathbf{a}^T \mathbf{D}_{DJ} \Sigma \mathbf{D}_{DJ}^T \mathbf{a}$ and $E \left[\hat{W}_{DF} \right] = E \left[\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_{DF} \hat{\Sigma}_k \mathbf{D}_{DF}^T \mathbf{a} \right] \rightarrow \mathbf{a}^T \mathbf{D}_{DJ} \Sigma \mathbf{D}_{DJ}^T \mathbf{a} + O(n^{-2})$. Under de-jure,

$$\hat{B}_{DJ} = \frac{1}{K-1} \sum_{k=1}^K \left[\pi_d (\bar{e}_{a,k} - \bar{e}_a) + \pi_d (\bar{\mathbf{P}}_{a,d} \mathbf{b}_{a,o,k} - \bar{\mathbf{P}}_{a,d} \bar{\mathbf{b}}_{a,o}) - \pi_{r,d} (\bar{e}_{r,k} - \bar{e}_r) - \pi_{r,d} (\bar{\mathbf{P}}_{r,d} \mathbf{b}_{r,o,k} - \bar{\mathbf{P}}_{r,d} \bar{\mathbf{b}}_{r,o}) \right]^2,$$

where $\bar{e}_{a,k} = \frac{1}{n_d} \sum_{i \in \mathcal{D}} e_{i,k}$, $\bar{e}_a = \frac{1}{K} \sum_{k=1}^K \bar{e}_{a,k}$, $\bar{\mathbf{P}}_{a,d} = \frac{1}{n_d} \sum_{i \in \mathcal{D}} \mathbf{P}_{a,d,i}$, $\bar{e}_{r,k} = \frac{1}{n_{r,d}} \sum_{i \in \mathcal{RD}} e_{i,k}$, $\bar{e}_r = \frac{1}{K} \sum_{k=1}^K \bar{e}_{r,k}$, $\bar{\mathbf{P}}_{r,d} = \frac{1}{n_{r,d}} \sum_{i \in \mathcal{RD}} \mathbf{P}_{r,d,i}$ and $\bar{\mathbf{b}}_{r,o} = \frac{1}{K} \sum_{k=1}^K \mathbf{b}_{r,o,k}$. This has expectation,

$$E \left[\hat{B}_{DJ} \right] = \pi_d^2 \left[\frac{\sigma^2 + n_d \bar{\mathbf{P}}_{a,d} \mathbf{V}_{a,o} \bar{\mathbf{P}}_{a,d}^T}{n_d} \right] + \pi_{r,d}^2 \left[\frac{\sigma^2 + n_{r,d} \bar{\mathbf{P}}_{r,d} \mathbf{V}_{r,o} \bar{\mathbf{P}}_{r,d}^T}{n_{r,d}} \right].$$

Under de-facto,

$$\hat{B}_{DF} = \frac{1}{K-1} \sum_{k=1}^K \left[\pi_d (\bar{e}_{a,k} - \bar{e}_a) + \pi_d (\bar{\mathbf{P}}_{a,d} \mathbf{b}_{DF,a,o,k} - \bar{\mathbf{P}}_{a,d} \bar{\mathbf{b}}_{DF,a,o}) - \pi_{r,d} (\bar{e}_{r,k} - \bar{e}_r) - \pi_{r,d} (\bar{\mathbf{P}}_{r,d} \mathbf{b}_{DF,r,o,k} - \bar{\mathbf{P}}_{r,d} \bar{\mathbf{b}}_{DF,r,o}) \right]^2,$$

where $\bar{\mathbf{b}}_{DF,a,o} = \frac{1}{K} \sum_{k=1}^K \mathbf{b}_{DF,a,o,k}$ and $\bar{\mathbf{b}}_{DF,r,o} = \frac{1}{K} \sum_{k=1}^K \mathbf{b}_{DF,r,o,k}$. This has expectation,

$$E \left[\hat{B}_{DF} \right] = \pi_d^2 \frac{\sigma^2}{n_d} + \pi_{r,d}^2 \frac{\sigma^2}{n_{r,d}} + \pi_d^2 \bar{\mathbf{P}}_{a,d} \mathbf{V}_{DF,a,o} \bar{\mathbf{P}}_{a,d}^T + \pi_{r,d}^2 \bar{\mathbf{P}}_{r,d} \mathbf{V}_{DF,r,o} \bar{\mathbf{P}}_{r,d}^T - 2\pi_d \pi_{r,d} \bar{\mathbf{P}}_{a,d} \text{Cov}(\mathbf{b}_{DF,a,o,k}, \mathbf{b}_{DF,r,o,k}) \bar{\mathbf{P}}_{r,d}^T.$$

Which can also be expressed as,

$$E \left[\hat{B}_{DF} \right] = E \left[\hat{B}_{DJ} \right] + \pi_d^2 \bar{\mathbf{P}}_{a,d} [\mathbf{V}_{DF,a,o} - \mathbf{V}_{a,o}] \bar{\mathbf{P}}_{a,d}^T + \pi_{r,d}^2 \bar{\mathbf{P}}_{r,d} [\mathbf{V}_{DF,r,o} - \mathbf{V}_{r,o}] \bar{\mathbf{P}}_{r,d}^T - 2\pi_d \pi_{r,d} \bar{\mathbf{P}}_{a,d} \text{Cov}(\mathbf{b}_{DF,a,o,k}, \mathbf{b}_{DF,r,o,k}) \bar{\mathbf{P}}_{r,d}^T.$$

The information anchored variance is,

$$E[V_{\text{anchored}}] = \mathbf{a}^T \mathbf{D}_{DJ} \Sigma \mathbf{D}_{DJ}^T \mathbf{a} + O(n^{-2}) + E \left[\hat{B}_{DJ} \right] + \frac{E \left[\hat{B}_{DJ} \right]}{E \left[\hat{W}_{DJ} \right]} O(n^{-2}).$$

If Rubin's variance estimator is information anchoring and preserves the information loss seen in the primary analysis under MAR then since $E \left[\hat{W}_{DJ} \right] = E[V_{DJ, \text{full}}]$,

$$0 \approx E \left[\hat{B}_{DJ} \right] - E \left[\hat{B}_{DF} \right] + \frac{E \left[\hat{B}_{DJ} \right]}{E \left[\hat{W}_{DJ} \right]} [O(n^{-2})]. \quad (4.10)$$

After substituting in the current results and simplifying this becomes,

$$0 \approx \pi_d^2 \bar{\mathbf{P}}_{a,d} [\mathbf{V}_{a,o} - \mathbf{V}_{DF,a,o}] \bar{\mathbf{P}}_{a,d}^T + \pi_{r,d}^2 \bar{\mathbf{P}}_{r,d} [\mathbf{V}_{r,o} - \mathbf{V}_{DF,r,o}] \bar{\mathbf{P}}_{r,d}^T + 2\pi_d \pi_{r,d} \bar{\mathbf{P}}_{a,d} \text{Cov}(\mathbf{b}_{DF,a,o,k}, \mathbf{b}_{DF,r,o,k}) \bar{\mathbf{P}}_{r,d}^T + \frac{E \left[\hat{B}_{DJ} \right]}{E \left[\hat{W}_{DJ} \right]} [O(n^{-2})]. \quad (4.11)$$

With deviation in both arms the difference between Rubin's variance estimator and the information anchored variance remains a small quantity in comparison to the information anchored variance.

An additional component which depends on the difference between the variance of the imputation parameters in the de-jure imputation model and de-facto imputation model for the reference arm, multiplied by $\pi_{r,d}^2$, is now included. The exact size of this additional piece depends on the specific de-facto scenario. But it will be similar in size to the first component on the RHS of (4.11), which we have already discussed is of a small order for realistic proportions of missing data. The covariance between the parameters of the active and reference arm de-facto imputation models also contributes to the sharpness of the approximation between Rubin's variance estimator and the information anchored variance. This is intuitive because the imputations in the arms may now be correlated. The exact size of this additional piece again depends on the specific de-facto scenario. But it is always multiplied by $\pi_d \pi_{r,d}$ thus will be of a relatively small order in practice. We now explore the implications for each de-facto option.

Under CR imputation, patients already in the reference arm will experience on-treatment MAR

behaviour. The imputation model for the deviating patients in the reference arm remains the same as under de-jure imputation, hence $\mathbf{V}_{DF,r,o} = \mathbf{V}_{r,o}$. The de-jure imputation model for the reference patients also becomes the imputation model for the deviating active patients, that is $\mathbf{V}_{DF,a,o} = \mathbf{V}_{r,o}$. Therefore in the CR setting result (4.11) is,

$$0 \approx \pi_d^2 \bar{\mathbf{P}}_{a,d} [\mathbf{V}_{a,o} - \mathbf{V}_{r,o}] \bar{\mathbf{P}}_{a,d}^T + 2\pi_d \pi_{r,d} \bar{\mathbf{P}}_{a,d} \mathbf{V}_{r,o} \bar{\mathbf{P}}_{r,d}^T + \frac{E[\hat{B}_{DJ}]}{E[\hat{W}_{DJ}]} [O(n^{-2})].$$

With deviation in both arms under CR, the sharpness of the approximation depends on the difference between $\mathbf{V}_{a,o}$ and $\mathbf{V}_{r,o}$, as well as the exact size of $\mathbf{V}_{r,o}$. When the variance of the parameters in the de-jure imputation model for the active patients matches the variance of the parameters in the de-jure imputation model for the reference patients, and both are small, improved information anchoring will be obtained.

Under LMCF imputation the covariance between the imputation parameters of the active and reference de-facto imputation models is zero. Hence in the LMCF setting result (4.11) is,

$$0 \approx \pi_d^2 \bar{\mathbf{P}}_{a,d} [\mathbf{V}_{a,o} - \mathbf{V}_{DF,a,o}] \bar{\mathbf{P}}_{a,d}^T + \pi_{r,d}^2 \bar{\mathbf{P}}_{r,d} [\mathbf{V}_{r,o} - \mathbf{V}_{DF,r,o}] \bar{\mathbf{P}}_{r,d}^T + \frac{E[\hat{B}_{DJ}]}{E[\hat{W}_{DJ}]} [O(n^{-2})].$$

With deviation in both arms under LMCF, the sharpness of the approximation depends on the difference between $\mathbf{V}_{a,o}$ and $\mathbf{V}_{DF,a,o}$, as well as the difference between $\mathbf{V}_{r,o}$ and $\mathbf{V}_{DF,r,o}$. That is when the variance of the parameters in the de-jure imputation model for the active patients matches the variance of the parameters in LMCF imputation model for the active patients and the variance of the parameters in the de-jure imputation model for the reference patients matches the variance of the parameters in LMCF imputation model for the reference patients, improved information anchoring will be obtained.

Under J2R or CIR imputation, the de-facto imputation model for patients in the reference arm is the same as under de-jure imputation, i.e. $\mathbf{V}_{DF,r,o} = \mathbf{V}_{r,o}$. Result (4.11) in the J2R/CIR settings becomes,

$$0 \approx \pi_d^2 \bar{\mathbf{P}}_{a,d} [\mathbf{V}_{a,o} - \mathbf{V}_{DF,a,o}] \bar{\mathbf{P}}_{a,d}^T + 2\pi_d \pi_{r,d} \bar{\mathbf{P}}_{a,d} \text{Cov}(\mathbf{b}_{DF,a,o,k}, \mathbf{b}_{r,o,k}) \bar{\mathbf{P}}_{r,d}^T + \frac{E[\hat{B}_{DJ}]}{E[\hat{W}_{DJ}]} [O(n^{-2})].$$

With deviation in both arms under CIR or J2R, the sharpness of the approximation depends on the difference between $\mathbf{V}_{a,o}$ and $\mathbf{V}_{DF,a,o}$. When the variance of the parameters in the de-

jure imputation model for the active patients ($\mathbf{V}_{a,o}$) matches the variance of the parameters in the de-facto imputation model ($\mathbf{V}_{DF,a,o}$), improved information anchoring will be obtained. The approximation will also be improved when the covariance between the parameters of the de-facto imputation model for the active patients and the parameters of the de-jure imputation model for the reference patients is small.

Implementation for improved information anchoring

With deviation in both arms the covariance between the parameters of the active and reference imputation models will additionally affect the sharpness of the approximation of the information anchored variance by Rubin’s variance estimator. Thus we could not simply rescale the variance parameters in the imputation models to obtain perfect information anchoring. We would need to also adjust for the covariance. The bootstrap approach, as outlined in Subsection 4.1.4, would also not achieve this goal. To remove the covariance terms, after bootstrapping the data to estimate the observed data distributions, we would need to draw separate data for each treatment arm to construct the imputation models.

We do not develop this approach further here, but highlight that the procedure described in Subsection 4.1.4 is not immediately applicable with deviation in both arms. Generally, in all de-facto settings (one of CR, J2R, CIR or LMCF) the approximation of the information anchored variance by Rubin’s variance estimator will be excellent, since the order of the terms in the difference between these quantities are relatively small in comparison to the information anchored variance itself. In Section 4.5 we explore the impact of deviation in both arms via simulation to demonstrate this.

4.2 Longitudinal setting with last measured variable subject to non-response

Here we present a general theory on the information anchoring properties of Rubin’s variance estimator in the longitudinal trial setting, where we have continuous measurement data recorded at baseline (time 1) and a further $J - 1$ follow-up visits and the last measured variable (at time J) is subject to non-response. Our focus is the treatment effect at the final time point. As in the previous section we begin by considering a two arm trial setting with deviation in one arm only. We later extend the proof for deviation in both treatment arms.

Here $V_{DJ, \text{full}}$ denotes the variance estimator for the treatment effect at the final time point in the design based analysis with fully observed data if no deviations occur; $V_{DF, \text{full}}$ denotes the variance estimator for the treatment effect at the final time point in the design based analysis where deviations occur off-treatment but the deviation data is fully observed (under one of CR, J2R, CIR or LMCF); $V_{DJ, \text{MI}}$ denotes Rubin’s variance estimator for the treatment effect at the final time point in the primary design based analysis where post-deviation data is unobserved but imputed under de-jure/MAR and $V_{DF, \text{MI}}$ denotes Rubin’s variance estimator for the treatment effect at the final time point in the design based analysis where post-deviation data is unobserved but imputed under de-facto behaviour (one of CR, J2R, CIR or LMCF).

4.2.1 Setting for Proposition 2

Consider a two arm trial consisting of $2n$ patients. n patients have been randomised to an active arm and n patients to a reference arm with a continuous outcome measured repeatedly over a series of visits, $j = 1, \dots, J$ (including baseline). The continuous outcome data at the final visit time J is contained in the $(2n \times 1)$ vector \mathbf{Y} . A corresponding $(2n \times p)$ matrix \mathbf{x} consists of the observed baseline covariates including baseline outcome and treatment, such that the estimated treatment effect can be written $\mathbf{a}^T \mathbf{Y}$ where \mathbf{a} may or may not incorporate \mathbf{x} .

We assume \mathbf{Y} is normally distributed and has known variance Σ . As before, the variance of the data is assumed to be known without any serious loss of generality, following inference by Carpenter and Kenward [44].

Suppose all n reference patients are observed at baseline and up to time J without deviation. However amongst the n active patients, only n_o patients are observed without deviation at all time points. The remaining n_d active arm patients deviate following time $J - 1$. For simplicity we assume there is no interim missing data. Let \mathcal{D} and \mathcal{O} define the sets of indices for the patients who do and do not deviate in the active arm respectively. Further suppose the variance-covariance matrix of baseline and follow-up is the same in both treatment groups. Interest lies in the treatment effect at time J .

Assume $V_{\text{DF, full}} = V_{\text{DJ, full}} + O(n^{-2})$. Typically, although not necessarily, some form of regression model will be the substantive model of interest, used to make inferences about the treatment effect in the population of interest. Appendix D.1.1 shows (4.2) holds for the mean treatment group difference, adjusted and unadjusted for baseline outcome in a trial with baseline (time 1) and $J - 1$ follow-up outcomes with non-response at time J .

4.2.2 Proposition 2

For sensitivity analyses of a two arm trial with baseline and $J - 1$ follow-up time points with non-response at time J , performed by MI using the algorithm given by Carpenter, Roger and Kenward [13], the difference between Rubin's variance estimator and the information anchored variance for the treatment effect at time J is,

$$0 \approx \pi_d^2 \bar{\mathbf{P}}_{a,d} (\mathbf{V}_{a,o} - \mathbf{V}_{DF,o}) \bar{\mathbf{P}}_{a,d}^T + \frac{E \left[\hat{B}_{\text{DJ}} \right]}{E \left[\hat{W}_{\text{DJ}} \right]} \left[O(n^{-2}) \right].$$

$\mathbf{V}_{a,o}$ represents the variance-covariance matrix of the parameter estimates in the de-jure imputation model and $\mathbf{V}_{DF,o}$ represents the variance-covariance matrix of the parameter estimates in the de-facto imputation model. $\bar{\mathbf{P}}_{a,d}$ is a vector containing the mean of the baseline covariates for each deviating active patient.

4.2.3 Proof of Proposition 2

Let $z = a, r$ index the randomised active arm or reference arm allocation for each patient i with follow-up outcome at time J denoted by Y_{ziJ} . The outcome data at the final time point for the reference patients are contained in the vector $\mathbf{Y}_{rJ} = (Y_{r1J}, \dots, Y_{rnJ})^T$. The final visit outcome data for the non-deviating active patients are contained in the vector $\mathbf{Y}_{aJ,o} = \{Y_{aiJ}; i \in \mathcal{O}\}^T$.

We suppose that each deviating patient has two potential outcomes at time J : the one that would occur if they remain on active treatment post-deviation (de-jure) and the other that would occur off-treatment post-deviation (de-facto).

The potentially observable de-jure data for the n_d deviating patients at time J are contained in the vector $\mathbf{Y}_{aJ,DJ,d}$ and the alternative de-facto outcome data in the vector $\mathbf{Y}_{aJ,DF,d}$. Define $\mathbf{Y} = (\mathbf{Y}_{rJ}, \mathbf{Y}_{aJ,o}, \mathbf{Y}_{aJ,DJ,d}, \mathbf{Y}_{aJ,DF,d})^T$ as the collection of observed and potentially observable outcome data, which has dimensions $[(n + n_o + 2n_d) \times 1]$.

For each deviating patient we can only observe one of the potential outcomes, either de-jure or de-facto. Consider two $[(n + n_o + 2n_d) \times (n + n_o + 2n_d)]$ matrices, \mathbf{D}_{DJ} and \mathbf{D}_{DF} of 0's and 1's such that $\mathbf{D}_{DJ}\mathbf{Y}$ gives the $[(n + n_o + 2n_d) \times 1]$ de-jure data and $\mathbf{D}_{DF}\mathbf{Y}$ gives the $[(n + n_o + 2n_d) \times 1]$ de-facto data at time J .

Let \mathbf{a} be a $[(n + n_o + 2n_d) \times 1]$ vector such that $\mathbf{a}^T\mathbf{D}_{DJ}\mathbf{Y}$ returns the de-jure treatment estimate and $\mathbf{a}^T\mathbf{D}_{DF}\mathbf{Y}$ returns the de-facto treatment estimate. When the deviating patients experience de-jure behaviour post-deviation and are observed the expectation of the variance of the de-jure on-treatment estimand can be expressed as in (4.1). We consider settings where the expectation of the variance of our de-facto estimand can then be expressed as in (4.2).

We now suppose that post-deviation data are unobserved, i.e. the potentially observable de-facto and de-jure entries in \mathbf{Y} are missing for the n_d active patients. We alternatively impute these outcomes, using de-jure imputation and de-facto imputation. This gives K 'complete' data samples \mathbf{Y}_k , of size $[(n + n_o + 2n_d) \times 1]$.

For this we need appropriate imputation distributions for the missing data under each scenario, with suitable posteriors for the included parameters. Under the de-jure assumption our imputation model is formed from the regression of $\mathbf{Y}_{aJ,o}$ on $\mathbf{P}_{a,o}$ where $\mathbf{P}_{a,o}$ is the design matrix for the imputation model, which contains the values of the $1, \dots, J - 1$ outcomes and covariates included in the imputation model (excluding treatment), along with a vector of 1's to include an intercept in the model for the n_o observed active patients.

The parameter estimates for the de-jure imputation model for the n_d patients missing outcome J are found using $\hat{\beta}_{a,o} = (\mathbf{P}_{a,o}^T\mathbf{P}_{a,o})^{-1}\mathbf{P}_{a,o}^T\mathbf{Y}_{aJ,o}$ with assumed known covariance matrix $\mathbf{V}_{a,o} = (\mathbf{P}_{a,o}^T\mathbf{P}_{a,o})^{-1}\sigma^2$.

We assume the large sample posterior for the parameters of the de-jure imputation model, denoted as $\hat{\beta}_{DJ}$, is normal and centered on the ML estimator $\hat{\beta}_{a,o}$ with covariance matrix $\mathbf{V}_{a,o}$. That is,

$$\hat{\beta}_{DJ}|Y_{a,J,o} \sim N(\hat{\beta}_{a,o}; \mathbf{V}_{a,o}).$$

The de-jure imputation model for active patient i deviating following time $J - 1$ and imputation k can therefore be expressed as,

$$\tilde{Y}_{aiJ,k}|\mathbf{Y}_{aJ,o} = \mathbf{P}_{a,d,i} \left[\hat{\beta}_{a,o} + \mathbf{b}_{a,o,k} \right] + e_{i,k} \text{ for } i \in \mathcal{D},$$

where, $\mathbf{b}_{a,o,k} \sim N(0, \mathbf{V}_{a,o})$, $e_{i,k} \sim N(0, \sigma^2)$ and $\mathbf{P}_{a,d,i}$ contains the values of the $1, \dots, J-1$ outcomes and covariates included in the imputation model (plus a 1 for the intercept) for each deviating patient i , who deviates following time $J - 1$.

For de-facto imputation we assume the large sample posterior for the imputation parameters for the n_d patients missing outcome J , $\hat{\beta}_{DF}$, is normal and centered on the ML estimator $\hat{\beta}_{DF,o}$ with known covariance matrix $\mathbf{V}_{DF,o}$, that is,

$$\hat{\beta}_{DF}|\mathbf{Y}_{DF,J,o} \sim N(\hat{\beta}_{DF,o}; \mathbf{V}_{DF,o}),$$

where $\mathbf{Y}_{DF,J,o}$ consists of the relevant observed outcome data under the particular de-facto setting of interest. The de-facto imputation model for active patient i deviating following time $J - 1$ and imputation k can therefore be expressed as,

$$\check{Y}_{aiJ,k}|\mathbf{Y}_{DF,J,o} = \mathbf{P}_{a,d,i} \left[\hat{\beta}_{DF,o} + \mathbf{b}_{DF,o,k} \right] + e_{i,k} \text{ for } i \in \mathcal{D},$$

where, $\mathbf{b}_{DF,o,k} \sim N(0, \mathbf{V}_{DF,o})$, and $e_{i,k} \sim N(0, \sigma^2)$. Under the assumption of equal variance-covariance matrix of baseline and follow-up by treatment arm we consequently assume the same variance for the residuals in the de-jure and de-facto imputation models (σ^2). In Subsection 4.2.5 we explore the impact of relaxing this assumption. We are interested in imputation inference for, $\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_{DJ} \mathbf{Y}_k$ or $\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_{DF} \mathbf{Y}_k$.

Letting the number of imputations, $K \rightarrow \infty$, the variance of our MI treatment estimate as estimated by Rubin's rules is, $V_{DJ, MI} = \hat{W}_{DJ} + \hat{B}_{DJ}$ or $V_{DF, MI} = \hat{W}_{DF} + \hat{B}_{DF}$ where under the conditions required in the proposition, $E \left[\hat{W}_{DJ} \right] = E \left[\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_{DJ} \hat{\Sigma}_k \mathbf{D}_{DJ}^T \mathbf{a} \right] \rightarrow \mathbf{a}^T \mathbf{D}_{DJ} \Sigma \mathbf{D}_{DJ}^T \mathbf{a}$ and $E \left[\hat{W}_{DF} \right] = E \left[\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_{DF} \hat{\Sigma}_k \mathbf{D}_{DF}^T \mathbf{a} \right] \rightarrow \mathbf{a}^T \mathbf{D}_{DF} \Sigma \mathbf{D}_{DF}^T \mathbf{a} + O(n^{-2})$.

Under de-jure,

$$\hat{B}_{\text{DJ}} = \frac{1}{K-1} \sum_{k=1}^K [\pi_d (\bar{e}_k - \bar{e}) + \pi_d (\bar{\mathbf{P}}_{a,d} \mathbf{b}_{a,o,k} - \bar{\mathbf{P}}_{a,d} \bar{\mathbf{b}}_{a,o})]^2,$$

where $\bar{e}_k = \frac{1}{n_d} \sum_{i \in \mathcal{D}} e_{i,k}$, $\bar{e} = \frac{1}{K} \sum_{k=1}^K \bar{e}_k$, $\bar{\mathbf{P}}_{a,d} = \frac{1}{n_d} \sum_{i \in \mathcal{D}} \mathbf{P}_{a,d,i}$ and $\bar{\mathbf{b}}_{a,o} = \frac{1}{K} \sum_{k=1}^K \mathbf{b}_{a,o,k}$. Which has expectation,

$$E[\hat{B}_{\text{DJ}}] = \pi_d^2 \left[\frac{\sigma^2 + n_d \bar{\mathbf{P}}_{a,d} \mathbf{V}_{a,o} \bar{\mathbf{P}}_{a,d}^T}{n_d} \right].$$

Under de-facto,

$$\hat{B}_{\text{DF}} = \frac{1}{K-1} \sum_{k=1}^K [\pi_d (\bar{e}_k - \bar{e}) + \pi_d (\bar{\mathbf{P}}_{a,d} \mathbf{b}_{DF,o,k} - \bar{\mathbf{P}}_{a,d} \bar{\mathbf{b}}_{DF,o})]^2,$$

where $\bar{\mathbf{b}}_{DF,o} = \frac{1}{K} \sum_{k=1}^K \mathbf{b}_{DF,o,k}$. Which has expectation,

$$E[\hat{B}_{\text{DF}}] = \pi_d^2 \left[\frac{\sigma^2 + n_d \bar{\mathbf{P}}_{a,d} \mathbf{V}_{DF,o} \bar{\mathbf{P}}_{a,d}^T}{n_d} \right].$$

The information anchored variance is,

$$E[V_{\text{anchored}}] = \mathbf{a}^T \mathbf{D}_{\text{DJ}} \boldsymbol{\Sigma} \mathbf{D}_{\text{DJ}}^T \mathbf{a} + O(n^{-2}) + E[\hat{B}_{\text{DJ}}] + \frac{E[\hat{B}_{\text{DJ}}]}{E[\hat{W}_{\text{DJ}}]} O(n^{-2}).$$

If Rubin's variance estimator is information anchoring and preserves the information loss in the primary analysis under MAR (4.4) holds, which in this setting is,

$$0 \approx \pi_d^2 \bar{\mathbf{P}}_{a,d} (\mathbf{V}_{a,o} - \mathbf{V}_{DF,o}) \bar{\mathbf{P}}_{a,d}^T + \frac{E[\hat{B}_{\text{DJ}}]}{E[\hat{W}_{\text{DJ}}]} [O(n^{-2})]. \quad (4.12)$$

This gives the required result. □

The difference between Rubin’s variance estimator and the information anchored variance is a small quantity, under the described conditions. We see the result corresponds with that obtained in the baseline and single follow-up setting. $\bar{\mathbf{P}}_{a,d}$, $\mathbf{V}_{a,o}$ and $\mathbf{V}_{DF,o}$ will be of greater dimensions here, but the order of these terms will remain the same i.e. the overall magnitude of these terms will be very small. The additional follow-up data does not have a marked impact when only the last measured variable is subject to non-response and the last measured time point is the focus of analysis. Rubin’s variance estimator provides an excellent approximation of the information anchored variance.

4.2.4 Implementation for improved information anchoring

As in the baseline and single follow-up setting, if the variance of the parameters in the de-facto imputation model corresponds to the variance of the parameters in the de-jure imputation model ($\mathbf{V}_{a,o} = \mathbf{V}_{DF,o}$) then Rubin’s variance estimator will better preserve the loss of information in the primary design based analysis in the current longitudinal trial setting with deviation in one arm. The first term on the RHS of (4.12) will disappear.

As described in Subsection 4.1.4 when this is not the case we can alter the reference based MI procedure by bootstrapping the observed reference sample, then drawing the required sample size from the estimated reference distribution to achieve $\mathbf{V}_{a,o} = \mathbf{V}_{DF,o}$. Alternatively we could re-scale the variance of the parameters in the de-facto imputation distribution to achieve $\mathbf{V}_{a,o} = \mathbf{V}_{DF,o}$. However as discussed in Subsection 4.1.3 the first term on the RHS of (4.12) is very small, relative to the information anchored variance we desire. Thus even without this additional step we will still see an excellent approximation between Rubin’s variance estimator and the desired information anchored variance.

4.2.5 Relaxing the equal variance assumption

As in Subsection 4.1.5 we can consider the impact of relaxing the assumption of equal variance-covariance matrix for baseline and follow-up across trial arms. Similar to the baseline and single follow-up setting when we relax the equal variance assumption we can no longer assume the variance of the residuals in the de-jure imputation model matches the variance of the residuals in the de-facto imputation model. In this case we denote the variance of the residuals in the de-jure imputation model as σ_{DJ}^2 and in the de-facto imputation model as σ_{DF}^2 . Then under de-facto imputation, the requirement for information anchoring becomes,

$$0 \approx \pi_d^2 \left[\frac{\sigma_{DJ}^2 - \sigma_{DF}^2}{n_d} + \bar{\mathbf{P}}_{a,d} (\mathbf{V}_{a,o} - \mathbf{V}_{DF,o}) \bar{\mathbf{P}}_{a,d}^T \right] + \frac{E \left[\hat{B}_{DJ} \right]}{E \left[\hat{W}_{DJ} \right]} [O(n^{-2})]. \quad (4.13)$$

Rubin’s variance estimator still approximates the information anchored variance estimator. The difference between Rubin’s variance estimator and the information anchored variance remains small in comparison to the information anchored variance. The additional component in the difference between Rubin’s variance and the information anchored variance is driven by the degree of the

difference in the variance structure by trial arm. But this additional component is multiplied by π_d/n i.e. is $O(\pi_d/n)$, thus will always be relatively small. The difference between $\mathbf{V}_{a,o} - \mathbf{V}_{DF,o}$ will also be slightly larger when the variance structures differ by trial arm. However as discussed in the single follow-up outcome setting in Subsection 4.1.5, the additional difference can be expressed as a component with the same order as $\mathbf{V}_{DF,o}$ in the equal variance setting and $\mathbf{V}_{a,o} - \mathbf{V}_{DF,o}$ remains multiplied by π_d^2 . Thus the order of this component will be the same as when variance-covariance structures were equal.

So we see that —similar to the baseline and single follow-up setting— relaxing the equal variance assumption does not greatly effect the approximation between Rubin’s variance estimator and the information anchored variance in the longitudinal setting, where the last measured variable is subject to non-response in one arm. Especially since we do not expect the variance structure to differ markedly by arm.

4.2.6 Extension for deviation in both arms

Following the approach outlined in Subsection 4.1.6 for the current longitudinal setting with deviation in both arms, we obtain a result which corresponds with that obtained in the baseline and single follow-up setting (4.11). The difference between Rubin’s variance estimator and information anchored variance is,

$$0 \approx \pi_d^2 \bar{\mathbf{P}}_{a,d} [\mathbf{V}_{a,o} - \mathbf{V}_{DF,a,o}] \bar{\mathbf{P}}_{a,d}^T + \pi_{r,d}^2 \bar{\mathbf{P}}_{r,d} [\mathbf{V}_{r,o} - \mathbf{V}_{DF,r,o}] \bar{\mathbf{P}}_{r,d}^T + 2\pi_d \pi_{r,d} \bar{\mathbf{P}}_{a,d} \text{Cov}(\mathbf{b}_{DF,a,o,k}, \mathbf{b}_{DF,r,o,k}) \bar{\mathbf{P}}_{r,d}^T + \frac{E[\hat{B}_{DJ}]}{E[\hat{W}_{DJ}]} [O(n^{-2})].$$

Rubin’s variance estimator still provides an excellent approximation of the information anchored variance. This is unsurprising given that the additional follow-up data did not have a marked impact when only the last measured variable was subject to non-response in one treatment arm.

4.3 Longitudinal setting with more than one deviation pattern

Often in a longitudinal trial where we have measurement data recorded at baseline (time 1) and $J - 1$ follow-up visits we will not only see non-response at the final response time but at earlier planned measurement occasions. Here we present a general theory on the information anchoring properties of Rubin’s variance estimator in the longitudinal trial setting, where visit times $j = 2, \dots, J$ are subject to non-response. Our focus is the treatment effect at the final time point. We restrict our attention to monotone missingness patterns, that is when response j is missing for a patient, responses $j + 1$ to J are also missing and responses $1, \dots, j - 1$ are observed. As in the previous sections we begin by focussing on the two arm trial setting with deviation in one arm only. Subsequently we extend the proof for deviation in both treatment arms.

4.3.1 Setting for Proposition 3

Consider a two arm trial consisting of $2n$ patients. n patients have been randomised to an active arm and n patients to a reference arm with a continuous outcome measured repeatedly over a series of visits, $j = 1, \dots, J$ (including baseline). The continuous outcome data at the final visit time J is contained in the $(2n \times 1)$ vector \mathbf{Y} . A corresponding $(2n \times p)$ matrix \mathbf{x} consists of the observed baseline covariates including baseline outcome and treatment, such that the estimated treatment effect can be written $\mathbf{a}^T \mathbf{Y}$. \mathbf{a} may or may not incorporate \mathbf{x} .

We assume \mathbf{Y} is normally distributed and has known variance Σ . The variance of the data is assumed to be known without any serious loss of generality, following inference by Carpenter and Kenward [44].

Suppose all n reference patients are observed at baseline and up to time J without deviation. However among the n active treatment patients, only n_o patients are observed without deviation at all time points. Among the deviating n_d active arm patients, we observe $n_{d,j}$ patients who deviate following time j , for $j = 1, \dots, J - 1$. For simplicity we assume there is no interim missing data. Let \mathcal{D} and \mathcal{O} define the sets of indices for the patients who do and do not deviate in the active arm respectively. Further \mathcal{DJ} denotes the set of indices for deviating patients who deviate following time j , so that $n_d = \sum_{j=1}^{J-1} n_{d,j}$. Let $\pi_{d,j} = n_{d,j}/n$. Assume the variance-covariance matrix of baseline and follow-up, is the same in both treatment groups. Interest lies in the treatment effect at time J .

Assume $V_{\text{DF, full}} = V_{\text{DJ, full}} + O(n^{-2})$. Appendix D.1.2 shows (4.2) holds for the mean treatment group difference (unadjusted and adjusted for baseline) in a trial with baseline (time 1) and $J - 1$ follow-up outcomes with monotone missingness patterns following baseline.

4.3.2 Proposition 3

For sensitivity analyses of a two arm longitudinal trial with monotone missingness performed by MI using the algorithm given by Carpenter, Roger and Kenward [13], the difference between Rubin's variance estimator and the information anchoring variance for the treatment effect at time J is,

$$0 \approx \sum_{j=1}^{(J-1)} \pi_{d,j}^2 \bar{\mathbf{P}}_{a,d,j} (\mathbf{V}_{a,o,j} - \mathbf{V}_{DF,o,j}) \bar{\mathbf{P}}_{a,d,j}^T + \frac{E[\hat{B}_{\text{DJ}}]}{E[\hat{W}_{\text{DJ}}]} [O(n^{-2})]. \quad (4.14)$$

$\mathbf{V}_{a,o,j}$ represents the variance-covariance matrix of the parameter estimates in the de-jure imputation model for patients deviating following time j and $\mathbf{V}_{DF,o,j}$ represents the variance-covariance matrix of the parameter estimates in the de-facto imputation model for patients deviating following time j . $\bar{\mathbf{P}}_{a,d,j}$ is a vector containing the mean of the baseline covariates for each deviating active patient who deviates following time j .

4.3.3 Proof of Proposition 3

Let $z = a, r$ index the randomised active arm or reference arm allocation for each patient i with follow-up outcome at time J denoted by Y_{ziJ} . The follow-up outcome data at the final time point for the reference patients are contained in the vector $\mathbf{Y}_{rJ} = (Y_{r1J}, \dots, Y_{rnJ})^T$. The final visit outcome data for the non-deviating active patients are contained in the vector $\mathbf{Y}_{aJ,o} = \{Y_{aiJ}; i \in \mathcal{O}\}^T$.

We suppose that each deviating patient has two potential outcomes at time J : the one that would occur if they remain on active treatment post-deviation (de-jure) and the other that would occur off-treatment post-deviation (de-facto).

The potentially observable de-jure data for the n_d deviating patients at time J are contained in the vector $\mathbf{Y}_{aJ,DJ,d}$ and the alternative de-facto outcome data in the vector $\mathbf{Y}_{aJ,DF,d}$. Define $\mathbf{Y} = (\mathbf{Y}_{rJ}, \mathbf{Y}_{aJ,o}, \mathbf{Y}_{aJ,DJ,d}, \mathbf{Y}_{aJ,DF,d})^T$ as the collection of observed and potentially observable outcome data, which has dimensions $[(n + n_o + 2n_d) \times 1]$.

For each deviating patient we can only observe one of the potential outcomes, either de-jure or de-facto. Consider two $[(n + n_o + 2n_d) \times (n + n_o + 2n_d)]$ matrices, \mathbf{D}_{DJ} and \mathbf{D}_{DF} of 0's and 1's such that $\mathbf{D}_{DJ}\mathbf{Y}$ gives the $[(n + n_o + 2n_d) \times 1]$ de-jure data and $\mathbf{D}_{DF}\mathbf{Y}$ gives the $[(n + n_o + 2n_d) \times 1]$ de-facto data at time J .

Let \mathbf{a} be a $[(n + n_o + 2n_d) \times 1]$ vector such that $\mathbf{a}^T\mathbf{D}_{DJ}\mathbf{Y}$ returns the de-jure treatment estimate and $\mathbf{a}^T\mathbf{D}_{DF}\mathbf{Y}$ returns the de-facto treatment estimate. When the deviating patients experience de-jure behaviour post-deviation and are observed the expectation of the variance of the de-jure on-treatment estimand can be expressed as in (4.1). We consider settings where the expectation of the variance of our de-facto estimand can then be expressed as in (4.2).

We now suppose that post-deviation data are unobserved, i.e. the potentially observable de-facto and de-jure entries in \mathbf{Y} are missing for the n_d active patients. We alternatively impute these outcomes, using de-jure imputation and de-facto imputation. This gives K 'complete' data samples \mathbf{Y}_k , of size $[(n + n_o + 2n_d) \times 1]$. For this we need appropriate imputation distributions for each missing data pattern under each scenario, with suitable posteriors for the included parameters.

Under our de-jure assumption (on-treatment MAR), the imputation model for patients deviating following time j , for $j = 1, \dots, J - 1$ is formed from the regression of $\mathbf{Y}_{aJ,o}$ on $\mathbf{P}_{a,o,j}$ where $\mathbf{P}_{a,o,j}$ is the design matrix for the imputation model, which contains the values of the $1, \dots, j$ outcomes and covariates included in the imputation model (excluding treatment) for the n_o observed active patients, along with a vector of 1's to include an intercept in the model. This is appropriate since we are not imputing any interim missing outcomes here. We only consider monotone missing data patterns. We are interested in the treatment effect at time J . As described by Carpenter and Kenward (p. 77–78) [44], under MAR, each of the regressions will be validly estimated from those observed in the data set.

The parameter estimates for the de-jure imputation model for the $n_{d,j}$ patients missing outcomes $j + 1$ to J are found using $\hat{\boldsymbol{\beta}}_{a,o,j} = (\mathbf{P}_{a,o,j}^T \mathbf{P}_{a,o,j})^{-1} \mathbf{P}_{a,o,j}^T \mathbf{Y}_{aJ,o}$ with assumed known covariance matrix $\mathbf{V}_{\mathbf{a},o,j} = (\mathbf{P}_{a,o,j}^T \mathbf{P}_{a,o,j})^{-1} \sigma_j^2$.

We assume the large sample posterior for the parameter estimates for the de-jure imputation model,

denoted $\hat{\beta}_{DJ,j}$, is normal and centered on the ML estimator $\hat{\beta}_{a,o,j}$ with covariance matrix $\mathbf{V}_{a,o,j}$. That is,

$$\hat{\beta}_{DJ,j}|Y_{aJ,o} \sim N(\hat{\beta}_{a,o,j}; \mathbf{V}_{a,o,j}).$$

The de-jure imputation model for active patient i deviating following time j , for $j = 1, \dots, J - 1$ and imputation k can therefore be expressed as,

$$\tilde{Y}_{aiJ,k}|Y_{aJ,o} = \mathbf{P}_{a,d,j,i} \left[\hat{\beta}_{a,o,j} + \mathbf{b}_{a,o,j,k} \right] + e_{i,j,k} \text{ for } i \in \{\mathcal{DJ}\},$$

where, $\mathbf{b}_{a,o,j,k} \sim N(0, \mathbf{V}_{a,o,j})$, $e_{i,j,k} \sim N(0, \sigma_j^2)$ and $\mathbf{P}_{a,d,j,i}$ contains the values of the $1, \dots, j$ outcomes and covariates included in the imputation model (excluding treatment, plus a 1 for the intercept) for each deviating active patient i , who deviates following time j .

For de-facto imputation we assume the large sample posterior for the imputation parameters for the $n_{d,j}$ patients missing outcomes $j + 1$ to J , $\hat{\beta}_{DF,j}$ is normal and centered on the ML estimator $\hat{\beta}_{DF,o,j}$ with known covariance matrix $\mathbf{V}_{DF,o,j}$, that is for $j = 1, \dots, J - 1$,

$$\hat{\beta}_{DF,j}|Y_{DF,J,o} \sim N(\hat{\beta}_{DF,o,j}; \mathbf{V}_{DF,o,j}),$$

where $\mathbf{Y}_{DF,J,o}$ consists of the relevant observed outcome data under the particular de-facto setting of interest. The de-facto imputation model for active patient i deviating following time j , for $j = 1, \dots, J - 1$ and imputation k can therefore be expressed as,

$$\check{Y}_{aiJ,k}|Y_{DF,J,o} = \mathbf{P}_{a,d,j,i} \left[\hat{\beta}_{DF,o,j} + \mathbf{b}_{DF,o,j,k} \right] + e_{i,j,k} \text{ for } i \in \{\mathcal{DJ}\},$$

where, $\mathbf{b}_{DF,o,j,k} \sim N(0, \mathbf{V}_{DF,o,j})$ and $e_{i,j,k} \sim N(0, \sigma_j^2)$. Under the assumption of equal variance-covariance matrix of baseline and follow-up by treatment arm we consequently assume the same variance for the residuals in the de-jure and de-facto imputation models for patients deviating following the same time j , for $j = 1, \dots, J - 1$. In Subsection 4.3.5 we consider the impact of relaxing this assumption. We are interested in imputation inference for, $\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_{DJ} \mathbf{Y}_k$ or $\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_{DF} \mathbf{Y}_k$.

Letting the number of imputations, $K \rightarrow \infty$, the variance of our MI treatment estimate as estimated by Rubin's rules is, $V_{DJ, MI} = \hat{W}_{DJ} + \hat{B}_{DJ}$ or $V_{DF, MI} = \hat{W}_{DF} + \hat{B}_{DF}$ where under the conditions required in the proposition, $E \left[\hat{W}_{DJ} \right] = E \left[\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_{DJ} \hat{\Sigma}_k \mathbf{D}_{DJ}^T \mathbf{a} \right] \rightarrow \mathbf{a}^T \mathbf{D}_{DJ} \Sigma \mathbf{D}_{DJ}^T \mathbf{a}$ and $E \left[\hat{W}_{DF} \right] = E \left[\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_{DF} \hat{\Sigma}_k \mathbf{D}_{DF}^T \mathbf{a} \right] \rightarrow \mathbf{a}^T \mathbf{D}_{DF} \Sigma \mathbf{D}_{DF}^T \mathbf{a} + O(n^{-2})$.

Under de-jure,

$$\hat{B}_{DJ} = \frac{1}{K-1} \sum_{k=1}^K \left[\sum_{j=1}^{(J-1)} \pi_{d,j} (\bar{e}_{j,k} - \bar{e}_j) + \pi_{d,j} (\bar{\mathbf{P}}_{a,d,j} \mathbf{b}_{a,o,j,k} - \bar{\mathbf{P}}_{a,d,j} \bar{\mathbf{b}}_{a,o,j}) \right]^2,$$

where $\bar{e}_{j,k} = \frac{1}{n_{d,j}} \sum_{i \in \mathcal{D}j} e_{i,j,k}$, $\bar{e}_j = \frac{1}{K} \sum_{k=1}^K \bar{e}_{j,k}$, $\bar{\mathbf{P}}_{a,d,j} = \frac{1}{n_{d,j}} \sum_{i \in \mathcal{D}j} \mathbf{P}_{a,d,j,i}$ and $\bar{\mathbf{b}}_{a,o,j} = \frac{1}{K} \sum_{k=1}^K \mathbf{b}_{a,o,j,k}$. Which has expectation,

$$E \left[\hat{B}_{DJ} \right] = \sum_{j=1}^{(J-1)} \pi_{d,j}^2 \left[\frac{\sigma_j^2 + n_{d,j} \bar{\mathbf{P}}_{a,d,j} \mathbf{V}_{a,o,j} \bar{\mathbf{P}}_{a,d,j}^T}{n_{d,j}} \right].$$

Under de-facto,

$$\hat{B}_{DF} = \frac{1}{K-1} \sum_{k=1}^K \left[\sum_{j=1}^{(J-1)} \pi_{d,j} (\bar{e}_{j,k} - \bar{e}_j) + \pi_{d,j} (\bar{\mathbf{P}}_{a,d,j} \mathbf{b}_{DF,o,j,k} - \bar{\mathbf{P}}_{a,d,j} \bar{\mathbf{b}}_{DF,o,j}) \right]^2,$$

where $\bar{\mathbf{b}}_{DF,o,j} = \frac{1}{K} \sum_{k=1}^K \mathbf{b}_{DF,o,j,k}$. Which has expectation,

$$E \left[\hat{B}_{DF} \right] = \sum_{j=1}^{(J-1)} \pi_{d,j}^2 \left[\frac{\sigma_j^2 + n_{d,j} \bar{\mathbf{P}}_{a,d,j} \mathbf{V}_{DF,o,j} \bar{\mathbf{P}}_{a,d,j}^T}{n_{d,j}} \right].$$

The information anchored variance is,

$$E [V_{\text{anchored}}] = \mathbf{a}^T \mathbf{D}_{DJ} \Sigma \mathbf{D}_{DJ}^T \mathbf{a} + O(n^{-2}) + E \left[\hat{B}_{DJ} \right] + \frac{E \left[\hat{B}_{DJ} \right]}{E \left[\hat{W}_{DJ} \right]} O(n^{-2}).$$

As in the previous settings, if Rubin's rules are information anchoring and preserve the information loss in the primary analysis under MAR (4.4) holds. Which in this setting is,

$$0 \approx \sum_{j=1}^{(J-1)} \left[\pi_{d,j}^2 \bar{\mathbf{P}}_{a,d,j} (\mathbf{V}_{a,o,j} - \mathbf{V}_{DF,o,j}) \bar{\mathbf{P}}_{a,d,j}^T \right] + \frac{E \left[\hat{B}_{DJ} \right]}{E \left[\hat{W}_{DJ} \right]} [O(n^{-2})] \quad (4.15)$$

This gives the required result in the longitudinal trial setting with monotone missingness in one treatment arm.

□

The result is a natural extension of that observed for the longitudinal setting with only one pattern of non-response (missingness at the final time point J). The approximation of the information anchored variance by Rubin's variance estimator is sharpened when, for each missing data pattern, the variance of the parameters in the de-facto imputation model matches the variance of the parameters in the de-jure (MAR) imputation model. However in practice the approximation is generally excellent regardless of an exact match in these quantities. Since this term and the other components in the difference will be of a smaller magnitude, relative to the information anchored variance, for realistic proportions of missing data.

4.3.4 Implementation for improved information anchoring

If the variance of the parameters in the de-jure imputation model corresponds to the variance of the parameters in the de-facto imputation model for each missing data pattern (i.e. $\mathbf{V}_{a,o,j} = \mathbf{V}_{DF,o,j}$ for $j = 1, \dots, J - 1$), Rubin's variance estimator will approximate the information anchored variance that preserves the loss of information in the primary design based analysis more closely in the longitudinal trial setting with monotone missingness patterns. The first term on the RHS of (4.15) will disappear for $j = 1, \dots, J - 1$.

As described in Subsection 4.1.4 when this is not the case we can alter the reference based procedure to achieve improved information anchoring via bootstrapping the observed reference case sample, then drawing the required samples from the estimated reference distribution to construct the imputed models to achieve $\mathbf{V}_{a,o,j} = \mathbf{V}_{DF,o,j}$ for $j = 1, \dots, J - 1$. Alternatively we could re-scale the variance of the parameters in the de-facto imputation model to ensure the variance corresponds with the variance of the parameters in the de-jure imputation model. With a large number of missing data patterns the re-scaling will become less trivial. Thus the bootstrap approach may be more desirable for trials with a larger number of missing data patterns. The choice of the approach undertaken is entirely at the preference of the trialist. However we note that the first term on the RHS of (4.12) is generally very small, relative to the information anchored variance. Thus without this condition we will still see an excellent approximation between Rubin's variance estimator and the desired information anchored variance.

4.3.5 Relaxing the equal variance assumption

When we relax the equal variance assumption we can no longer assume the variance of the residuals in the de-jure imputation model for patients with missingness pattern j matches the variance of the residuals in the de-facto imputation model for patients with missingness pattern j , for each missing data pattern j .

In this setting we denote the variance of the residuals in the de-jure imputation model for patients missing outcomes $j + 1, \dots, J$ as $\sigma_{DJ,j}^2$ and in the de-facto imputation model as $\sigma_{DF,j}^2$ for $j = 1, \dots, J - 1$. Then under de-jure imputation, the requirement for information anchoring becomes,

$$0 \approx \sum_{j=1}^{(J-1)} \pi_{d,j}^2 \left[\frac{\sigma_{DJ,j}^2 - \sigma_{DF,j}^2}{n_{d,j}} + \bar{\mathbf{P}}_{a,d,j} (\mathbf{V}_{a,o,j} - \mathbf{V}_{DF,o,j}) \bar{\mathbf{P}}_{a,d,j}^T \right] + \frac{E[\hat{B}_{DJ}]}{E[\hat{W}_{DJ}]} [O(n^{-2})]. \quad (4.16)$$

For each missingness pattern with deviation following time j , an additional component is incorporated. The new components in the difference between Rubin's variance and the ideal information anchored variance are driven by the degree of difference in the variance structure by trial arm for each missingness pattern. Since the variance structure is not likely to differ too markedly by trial arm for each missingness pattern, and these extra components are each multiplied by $\pi_{d,j}/n$, the overall impact will in practice be relatively small.

So we see that similar to the longitudinal setting where the last measured variable is subject to non-response, relaxing the equal variance assumption does not greatly effect the approximation between Rubin's variance estimator and the ideal information anchored variance in the longitudinal setting with monotone non-response in one arm.

4.3.6 Extension for deviation in both arms

Suppose among the n reference patients only $n_{r,o}$ are actually observed at all time points without deviation. Among the remaining $n_{r,d}$ deviating reference patients, we observe $n_{r,d,j}$ patients who deviate following time j for $j = 1, \dots, J-1$. For simplicity we assume there is no interim missing data in the reference arm. Let \mathcal{RD} and \mathcal{RO} define the sets of indices for patients who do and do not deviate in the reference arm respectively. Further $\mathcal{RD}\mathcal{J}$ denotes the set of indices for deviating reference patients who deviate following time j , so that $n_{r,d} = \sum_{j=1}^{J-1} n_{r,d,j}$. Interest still lies in the treatment effect at time J .

The outcome data for the observed reference patients at the final time point are contained in the vector $\mathbf{Y}_{rJ,o} = \{Y_{riJ}; i \in \mathcal{RO}\}^T$. The potentially observable de-jure data for the $n_{r,d}$ deviating reference patients are contained in the vector $\mathbf{Y}_{rJ,DJ,d}$ and the alternative de-facto outcome data in the vector $\mathbf{Y}_{rJ,DF,d}$. The full collection of observed and potentially observable outcome data is now defined as $\mathbf{Y} = (\mathbf{Y}_{rJ,o}, \mathbf{Y}_{rJ,DJ,d}, \mathbf{Y}_{rJ,DF,d}, \mathbf{Y}_{aJ,o}, \mathbf{Y}_{aJ,DJ,d}, \mathbf{Y}_{aJ,DF,d})^T$ which has dimensions $[(n_{r,o} + 2n_{r,d} + n_o + 2n_d)]$. We assume \mathbf{Y} is normally distributed and has known variance Σ .

We redefine the two $[(n_{r,o} + 2n_{r,d} + n_o + 2n_d) \times (n_{r,o} + 2n_{r,d} + n_o + 2n_d)]$ matrices \mathbf{D}_{DJ} and \mathbf{D}_{DF} of 0's and 1's so that $\mathbf{D}_{DJ}\mathbf{Y}$ and $\mathbf{D}_{DF}\mathbf{Y}$ now each give the $[(n_{r,o} + 2n_{r,d} + n_o + 2n_d) \times 1]$ de-jure data or $[(n_{r,o} + 2n_{r,d} + n_o + 2n_d) \times 1]$ de-facto data across both treatment arms. We focus on settings where $E[V_{DF, \text{full}}] = \mathbf{a}^T \mathbf{D}_{DJ} \Sigma \mathbf{D}_{DJ}^T \mathbf{a} + O(n^{-2})$.

We follow the steps outlined above to establish the de-jure imputation model for active arm patient i , deviating following time j for $j = 1, \dots, J-1$ and imputation k as,

$$\tilde{Y}_{aiJ,k} | \mathbf{Y}_{aJ,o} = \mathbf{P}_{a,d,j,i} \left[\hat{\beta}_{a,o,j} + \mathbf{b}_{a,o,j,k} \right] + e_{i,j,k} \text{ for } i \in \{\mathcal{DJ}\},$$

where $\mathbf{b}_{a,o,j,k} \sim N(0, \mathbf{V}_{a,o,j})$, $e_{i,j,k} \sim N(0, \sigma_j^2)$ and $\mathbf{P}_{a,d,j,i}$ contains the values of the $1, \dots, j$ outcomes and covariates included in the imputation model (excluding treatment, plus a 1 for the intercept) for each deviating active patient i , who deviates following time j . For the reference arm, under our de-jure (on-treatment MAR) assumption the imputation model for patients deviating following time j , for $j = 1, \dots, J - 1$ is formed from the regression of $\mathbf{Y}_{rJ,o}$ on $\mathbf{P}_{r,o,j}$ where $\mathbf{P}_{r,o,j}$ is the design matrix for the imputation model, which contains the values of the $1, \dots, j$ outcomes and covariates included in the imputation model with a vector of 1's to include an intercept term in the model, for the $n_{r,o}$ observed reference patients.

The parameter estimates for the de-jure reference arm imputation model for the $n_{r,d,j}$ patients missing outcomes $j + 1$ to J are found using $\hat{\boldsymbol{\beta}}_{r,o,j} = (\mathbf{P}_{r,o,j}^T \mathbf{P}_{r,o,j})^{-1} \mathbf{P}_{r,o,j}^T \mathbf{Y}_{rJ,o}$ with assumed known covariance matrix $\mathbf{V}_{r,o,j} = (\mathbf{P}_{r,o,j}^T \mathbf{P}_{r,o,j})^{-1} \sigma_j^2$. We assume the large sample posterior for the parameter estimates for the de-jure reference arm imputation model, denoted $\hat{\boldsymbol{\beta}}_{DJ,r,j}$, is normal and centered on the ML estimator $\hat{\boldsymbol{\beta}}_{r,o,j}$ with covariance matrix $\mathbf{V}_{r,o,j}$. That is,

$$\hat{\boldsymbol{\beta}}_{DJ,r,j} | \mathbf{Y}_{rJ,o} \sim N(\hat{\boldsymbol{\beta}}_{r,o,j}; \mathbf{V}_{r,o,j}).$$

The de-jure imputation model for reference patient i deviating following time j , for $j = 1, \dots, J - 1$ and imputation k can therefore be expressed as,

$$\tilde{Y}_{riJ,k} | \mathbf{Y}_{rJ,o} = \mathbf{P}_{r,d,j,i} \left[\hat{\boldsymbol{\beta}}_{r,o,j} + \mathbf{b}_{r,o,j,k} \right] + e_{i,j,k} \text{ for } i \in \{\mathcal{RDJ}\},$$

where $\mathbf{b}_{r,o,j,k} \sim N(0, \mathbf{V}_{r,o,j})$, $e_{i,j,k} \sim N(0, \sigma_j^2)$ and $\mathbf{P}_{r,d,j,i}$ contains the values of the $1, \dots, j$ outcomes and covariates included in the imputation model (excluding treatment, plus a 1 for the intercept) for each deviating reference patient i , who deviates following time j . Under de-facto imputation for patients in the active arm deviating following time j for $j = 1, \dots, J - 1$ we assume the large sample posterior for the parameters of the imputation model, which we denote by $\hat{\boldsymbol{\beta}}_{DF,a,j}$, is normal and centered on the ML estimator $\hat{\boldsymbol{\beta}}_{DF,a,o,j}$ with known covariance matrix $\mathbf{V}_{DF,a,o,j}$. That is for $j = 1, \dots, J - 1$,

$$\hat{\boldsymbol{\beta}}_{DF,a,j} | \mathbf{Y}_{DF,a,J,o} \sim N\left(\hat{\boldsymbol{\beta}}_{DF,a,o,j}, \mathbf{V}_{DF,a,o,j}\right),$$

where $\mathbf{Y}_{DF,a,J,o}$ consists of the relevant observed outcome data under the particular de-facto setting of interest. The de-facto imputation model for active patient i deviating following time j for $j = 1, \dots, J - 1$ and imputation k can therefore be expressed as,

$$\tilde{Y}_{aiJ,k} | \mathbf{Y}_{DF,a,J,o} = \mathbf{P}_{a,d,j,i} \left[\hat{\boldsymbol{\beta}}_{DF,a,o,j} + \mathbf{b}_{DF,a,o,j,k} \right] + e_{i,j,k} \text{ for } i \in \mathcal{DJ},$$

where $\mathbf{b}_{DF,a,o,j,k} \sim N(0, \mathbf{V}_{DF,a,o,j})$, and $e_{i,j,k} \sim N(0, \sigma_j^2)$. Under de-facto imputation for patients in the reference arm deviating following time j for $j = 1, \dots, J-1$ we assume the large sample posterior for the parameters of the imputation model, which we denote by $\hat{\boldsymbol{\beta}}_{DF,r,j}$, is normal and centered on the ML estimator $\hat{\boldsymbol{\beta}}_{DF,r,o,j}$ with known covariance matrix $\mathbf{V}_{DF,r,o,j}$. That is for $j = 1, \dots, J-1$,

$$\hat{\boldsymbol{\beta}}_{DF,r,j} | \mathbf{Y}_{DF,r,J,o} \sim N\left(\hat{\boldsymbol{\beta}}_{DF,r,o,j}, \mathbf{V}_{DF,r,o,j}\right),$$

where $\mathbf{Y}_{DF,r,J,o}$ consists of the relevant observed outcome data under the particular de-facto setting of interest. The de-facto imputation model for reference patient i deviating following time j for $j = 1, \dots, J-1$ and imputation k can therefore be expressed as,

$$\tilde{Y}_{riJ,k} | \mathbf{Y}_{DF,r,J,o} = \mathbf{P}_{r,d,j,i} \left[\hat{\boldsymbol{\beta}}_{DF,r,o,j} + \mathbf{b}_{DF,r,o,j,k} \right] + e_{i,j,k} \text{ for } i \in \mathcal{RDJ},$$

where $\mathbf{b}_{DF,r,o,j,k} \sim N(0, \mathbf{V}_{DF,r,o,j})$, and $e_{i,j,k} \sim N(0, \sigma_j^2)$. We are interested in imputation inference for, $\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_{DJ} \mathbf{Y}_k$ or $\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_{DF} \mathbf{Y}_k$. For Rubin's variance estimator, under the described conditions, $E \left[\hat{W}_{DJ} \right] = E \left[\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_{DJ} \hat{\boldsymbol{\Sigma}}_k \mathbf{D}_{DJ}^T \mathbf{a} \right] \rightarrow \mathbf{a}^T \mathbf{D}_{DJ} \boldsymbol{\Sigma} \mathbf{D}_{DJ}^T \mathbf{a}$ and $E \left[\hat{W}_{DF} \right] = E \left[\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_{DF} \hat{\boldsymbol{\Sigma}}_k \mathbf{D}_{DF}^T \mathbf{a} \right] \rightarrow \mathbf{a}^T \mathbf{D}_{DJ} \boldsymbol{\Sigma} \mathbf{D}_{DJ}^T \mathbf{a} + O(n^{-2})$. Under de-jure,

$$\begin{aligned} \hat{B}_{DJ} = \frac{1}{K-1} \sum_{k=1}^K \left[\sum_{j=1}^{(J-1)} \pi_{d,j} (\bar{e}_{a,j,k} - \bar{e}_{a,j}) + \pi_{d,j} (\bar{\mathbf{P}}_{a,d,j} \mathbf{b}_{a,o,j,k} - \bar{\mathbf{P}}_{a,d,j} \bar{\mathbf{b}}_{a,o,j}) \right. \\ \left. - \pi_{r,d,j} (\bar{e}_{r,j,k} - \bar{e}_{r,j}) - \pi_{r,d,j} (\bar{\mathbf{P}}_{r,d,j} \mathbf{b}_{r,o,j,k} - \bar{\mathbf{P}}_{r,d,j} \bar{\mathbf{b}}_{r,o,j}) \right]^2 \end{aligned}$$

where $\bar{e}_{a,j,k} = \frac{1}{n_{d,j}} \sum_{i \in \mathcal{DJ}} e_{i,j,k}$, $\bar{e}_{a,j} = \frac{1}{K} \sum_{k=1}^K \bar{e}_{a,j,k}$, $\bar{\mathbf{P}}_{a,d,j} = \frac{1}{n_{d,j}} \sum_{i \in \mathcal{DJ}} \mathbf{P}_{a,d,j,i}$, $\bar{\mathbf{b}}_{a,o,j} = \frac{1}{K} \sum_{k=1}^K \mathbf{b}_{a,o,j,k}$, $\bar{e}_{r,j,k} = \frac{1}{n_{d,j}} \sum_{i \in \mathcal{RDJ}} e_{i,j,k}$, $\bar{e}_{r,j} = \frac{1}{K} \sum_{k=1}^K \bar{e}_{r,j,k}$, $\bar{\mathbf{P}}_{r,d,j} = \frac{1}{n_{r,d,j}} \sum_{i \in \mathcal{RDJ}} \mathbf{P}_{r,d,j,i}$ and $\bar{\mathbf{b}}_{r,o,j} = \frac{1}{K} \sum_{k=1}^K \mathbf{b}_{r,o,j,k}$. This has expectation,

$$E \left[\hat{B}_{DJ} \right] = \sum_{j=1}^{(J-1)} \pi_{d,j}^2 \left[\frac{\sigma_j^2 + n_{d,j} \bar{\mathbf{P}}_{a,d,j} \mathbf{V}_{a,o,j} \bar{\mathbf{P}}_{a,d,j}^T}{n_{d,j}} \right] + \pi_{r,d,j}^2 \left[\frac{\sigma_j^2 + n_{r,d,j} \bar{\mathbf{P}}_{r,d,j} \mathbf{V}_{r,o,j} \bar{\mathbf{P}}_{r,d,j}^T}{n_{r,d,j}} \right].$$

Under de-facto,

$$\begin{aligned} \hat{B}_{DF} = \frac{1}{K-1} \sum_{k=1}^K \left[\sum_{j=1}^{(J-1)} \pi_{d,j} (\bar{e}_{a,j,k} - \bar{e}_{a,j}) + \pi_{d,j} (\bar{\mathbf{P}}_{a,d,j} \mathbf{b}_{DF,a,o,j,k} - \bar{\mathbf{P}}_{a,d,j} \bar{\mathbf{b}}_{DF,a,o,j}) \right. \\ \left. - \pi_{r,d,j} (\bar{e}_{r,j,k} - \bar{e}_{r,j}) - \pi_{r,d,j} (\bar{\mathbf{P}}_{r,d,j} \mathbf{b}_{DF,r,o,j,k} - \bar{\mathbf{P}}_{r,d,j} \bar{\mathbf{b}}_{DF,r,o,j}) \right]^2 \end{aligned}$$

where $\bar{\mathbf{b}}_{DF,a,o,j} = \frac{1}{K} \sum_{k=1}^K \mathbf{b}_{DF,a,o,j,k}$ and $\bar{\mathbf{b}}_{DF,r,o,j} = \frac{1}{K} \sum_{k=1}^K \mathbf{b}_{DF,r,o,j,k}$. Which has expectation,

$$E \left[\hat{B}_{DF} \right] = \sum_{j=1}^{(J-1)} \pi_{d,j}^2 \left[\frac{\sigma_j^2 + n_{d,j} \bar{\mathbf{P}}_{\mathbf{a},d,j} \mathbf{V}_{DF,a,o,j} \bar{\mathbf{P}}_{\mathbf{a},d,j}^T}{n_{d,j}} \right] + \pi_{r,d,j}^2 \left[\frac{\sigma_j^2 + n_{r,d,j} \bar{\mathbf{P}}_{r,d,j} \mathbf{V}_{DF,r,o,j} \bar{\mathbf{P}}_{r,d,j}^T}{n_{r,d,j}} \right] \\ - 2\pi_{d,j} \pi_{r,d,j} \bar{\mathbf{P}}_{\mathbf{a},d,j} \text{Cov}(\mathbf{b}_{DF,a,o,j,k}, \mathbf{b}_{DF,r,o,j,k}) \bar{\mathbf{P}}_{r,d,j}^T.$$

The information anchored variance is,

$$E[V_{\text{anchored}}] = \mathbf{a}^T \mathbf{D}_{DJ} \boldsymbol{\Sigma} \mathbf{D}_{DJ}^T \mathbf{a} + O(n^{-2}) + E \left[\hat{B}_{DJ} \right] + \frac{E \left[\hat{B}_{DJ} \right]}{E \left[\hat{W}_{DJ} \right]} O(n^{-2}).$$

If Rubin's variance estimator is information anchoring and preserves the information loss in the primary analysis under MAR then (4.4) holds. In this setting (4.4) becomes,

$$0 \approx \sum_{j=1}^{(J-1)} \pi_{d,j}^2 \bar{\mathbf{P}}_{\mathbf{a},d,j} [\mathbf{V}_{a,o,j} - \mathbf{V}_{DF,a,o,j}] \bar{\mathbf{P}}_{\mathbf{a},d,j}^T + \pi_{r,d,j}^2 \bar{\mathbf{P}}_{r,d,j} [\mathbf{V}_{r,o,j} - \mathbf{V}_{DF,r,o,j}] \bar{\mathbf{P}}_{r,d,j}^T \\ + 2\pi_{d,j} \pi_{r,d,j} \bar{\mathbf{P}}_{\mathbf{a},d,j} \text{Cov}(\mathbf{b}_{DF,a,o,j,k}, \mathbf{b}_{DF,r,o,j,k}) \bar{\mathbf{P}}_{r,d,j}^T + \frac{E \left[\hat{B}_{DJ} \right]}{E \left[\hat{W}_{DJ} \right]} [O(n^{-2})]. \quad (4.17)$$

With deviation in both arms of a longitudinal trial with monotone missingness, the difference between Rubin's variance estimator and the information anchored variance remains a small quantity. Result (4.17) extends the inference drawn in the longitudinal setting where only the last measured variable was subject to response. With $j = 1, \dots, J - 1$ missing data patterns the sharpness of the approximation depends on the degree of similarity between the variance of the de-jure imputation model parameters and de-facto imputation model parameters, for both the active arm and reference arm, for each missing data pattern. The magnitude of the covariance between the active and reference arm de-facto imputation model parameters for each missing data pattern also contributes to the sharpness of the approximation. In practice however these terms will be negligible in size in comparison to the information anchored variance. In Section 4.5 we conduct a simulation study to further illustrate this.

So far, throughout this chapter we have assumed an infinite number of imputations. We now explore the impact a finite number of imputations has on the information anchoring performance of Rubin's variance estimator.

4.4 Finite imputations

The theoretical results in this chapter have concentrated on the setting where the number of imputations, $K \rightarrow \infty$. In practice this will not be the case, however results will still hold for finite K . For finite K the variance of our MI treatment estimate as estimated by Rubin's rules is, $V_{\text{DJ, MI}} = \hat{W}_{\text{DJ}} + \left(1 + \frac{1}{K}\right) \hat{B}_{\text{DJ}}$ or $V_{\text{DF, MI}} = \hat{W}_{\text{DF}} + \left(1 + \frac{1}{K}\right) \hat{B}_{\text{DF}}$. We will therefore have additional terms in the difference between Rubin's variance estimator and the ideal information anchored variance, but these will also be very small. They will be the same order of the terms already presented multiplied by K^{-1} , thus smaller than the components discussed throughout this chapter.

To see this explicitly we re-consider the proof of Proposition 1 for finite K . If Rubin's variance estimator preserves the information loss seen in the primary analysis under MAR in the sensitivity analysis then,

$$E \left[\hat{W}_{\text{DF}} \right] + \left(1 + \frac{1}{K} \right) E \left[\hat{B}_{\text{DF}} \right] \approx E \left[V_{\text{DF, full}} \right] \left[1 + \frac{\left(1 + \frac{1}{K} \right) E \left[\hat{B}_{\text{DJ}} \right]}{E \left[\hat{W}_{\text{DJ}} \right]} \right].$$

That is,

$$\begin{aligned} & \mathbf{a}^T \mathbf{D}_{\text{DJ}} \Sigma \mathbf{D}_{\text{DJ}}^T \mathbf{a} + O(n^{-2}) + \left(1 + \frac{1}{K} \right) E \left[\hat{B}_{\text{DF}} \right] \\ & \approx \mathbf{a}^T \mathbf{D}_{\text{DJ}} \Sigma \mathbf{D}_{\text{DJ}}^T \mathbf{a} + O(n^{-2}) + \frac{\left(1 + \frac{1}{K} \right) E \left[\hat{B}_{\text{DJ}} \right]}{E \left[\hat{W}_{\text{DJ}} \right]} \left[\mathbf{a}^T \mathbf{D}_{\text{DJ}} \Sigma \mathbf{D}_{\text{DJ}}^T \mathbf{a} + O(n^{-2}) \right]. \end{aligned}$$

After simplification and rearrangement this becomes,

$$0 \approx \left(1 + \frac{1}{K} \right) \left[E \left[\hat{B}_{\text{DJ}} \right] - E \left[\hat{B}_{\text{DF}} \right] + \frac{E \left[\hat{B}_{\text{DJ}} \right]}{E \left[\hat{W}_{\text{DJ}} \right]} \left[O(n^{-2}) \right] \right].$$

Which is,

$$0 \approx \left(1 + \frac{1}{K} \right) \left[(\pi_d)^2 \bar{\mathbf{P}}_{a,d} (\mathbf{V}_{a,o} - \mathbf{V}_{DF,o}) \bar{\mathbf{P}}_{a,d}^T + \frac{E \left[\hat{B}_{\text{DJ}} \right]}{E \left[\hat{W}_{\text{DJ}} \right]} \left[O(n^{-2}) \right] \right].$$

We see sensitivity analyses performed by MI using the algorithm given by Carpenter, Roger and Kenward [13] are also information anchoring with finite K . We infer that Proposition 2 and 3

will similarly still hold. We will only have additional smaller terms in the approximation between Rubin’s variance estimator and the information anchored variance. To demonstrate this we now conduct a simulation study. In each MI scenario $K = 50$ imputations are used.

4.5 Simulation study to explore Rubin’s variance estimator in the longitudinal setting

This section reports the results of a simulation study to assess the information anchoring properties of Rubin’s variance formula in a variety of longitudinal trial settings following reference based MI, for each of CR, J2R, CIR and LMCF.

4.5.1 Methods

Data from the asthma trial formed the motivating example for generating the datasets. Baseline (time 1), time 2 and time 3 data were generated using an underlying MVN distribution with means and covariance matrices similar to those obtained in the asthma trial at baseline, week 4 and week 12 as,

$$\mu_{\text{placebo}} = [2, 1.95, 1.9], \mu_{\text{active}} = [2, 2.21, \mu_{a,3}],$$

$$\Sigma_{\text{placebo}} = \Sigma_{\text{active}} = \begin{bmatrix} 0.4 & 0.2 & 0.2 \\ 0.2 & 0.5 & 0.2 \\ 0.2 & 0.2 & 0.6 \end{bmatrix}.$$

In the study $\mu_{a,3} \approx 2.2$ corresponding to a treatment difference of $\Delta = 0.3$. We consider a range of $\mu_{a,3}$ corresponding to treatment differences of $\Delta = \{0, 0.3, 1\}$ at time 3 for a range of monotone MCAR missingness patterns among the active patients; with 5–35% of active patients missing data at time 2 and time 3 and a further 5–35% of active patients missing time 3 only, corresponding to a total of 10–70% missing data at time 3 in the active arm. Table E.1 summarises the specific proportions of missing data in the active arm at each time point in each scenario.

One thousand independent complete datasets were generated for each combination of $\mu_{a,3}$ and missingness, and a sample of $n=500$ patients ($n=250$ placebo, $n=250$ active) was drawn for all simulations. For each setting, the treatment effect and its estimated variance ($V_{\text{DJ, full}}$) were computed using the full data. The analysis model was a linear regression of the time 3 response on randomised treatment group.

Monotone MCAR missingness was achieved by independently sampling a binary indicator from a Bernoulli distribution with success probability equal to the proportion of missing data required at time 2 for patients in the active arm (half that required by time 3). If the indicator equalled one at time 2 then the corresponding outcomes at time 2 and 3 were set missing. A second binary indicator was independently sampled for active individuals observed at time 2 from a Bernoulli

distribution with success probability equal to the proportion of additional missing data required at time 3 to obtain the second missingness pattern. This resulted in a sample of n_o patients observed at all time point, $n_{d,1}$ patients observed with deviation following baseline (time 1) and $n_{d,2}$ patients with deviation following time 2 in the active arm.

The reference based MI approach of Carpenter, Roger and Kenward [13] was applied to each incomplete dataset under MAR, CR, J2R, CIR and LMCF with 50 imputations. In order to minimise the Monte-Carlo variability in our comparisons, we used the same set of 1000 incomplete datasets for each imputation procedure. In each de-facto scenario, and for each of the 1000 incomplete datasets, we created a full dataset by generating the post-deviation data under the scenario-specific assumptions using the appropriate conditional normal distributions. This enabled the treatment effect and its variance, had we been able to observe post-deviation data under the particular de-facto setting ($V_{DF, full}$), to be computed. The ideal information anchored variance computed as $(V_{DJ, MI}/V_{DJ, full}) \times V_{DF, full}$ and Rubin’s estimate of variance were assessed and compared. Estimates were averaged over the 1000 simulations in each scenario for comparison.

We then considered longitudinal trial settings with monotone deviation in both arms for the asthma setting with $\Delta = 0.3$; with 5–35% of patients missing data at time 2 and time 3 and a further 5–35% of patients missing data at time 3 only in both treatment arms, corresponding to total of 10–70% missingness in both arms at time 3. Table E.1 summarises the specific proportions of missing data imposed in both arms at each time point in each scenario. We also investigated the setting where the amount of deviation in the reference arm was half that observed for the active arm; with 5–35% of patients missing data at time 2 and time 3 and a further 5–35% of patients missing data at time 3 only in the active arm (total of 10–70% missingness at time 3 in the active arm) and 2.5%–17.5% of patients missing data at time 2 and a further 2.5–17.5% missing data at time 3 only in the reference arm (total of 5–35% missingness at time 2 in the reference arm). Table E.2 summarises the specific proportions of missing data imposed at each time point, in each scenario, by trial arm.

Additional simulations were run for the asthma inspired setting with $\Delta = 0.3$ and a smaller ($n=100$) and larger ($n=1000$) sample size per arm. We also considered the effect of having a smaller or larger covariance between each time point for the original mean structure ($\Delta = 0.3$) and sample size of $n=250$ per arm for each missingness scenario as follows,

Low covariance set-up,

$$\Sigma_{\text{placebo,low}} = \Sigma_{\text{active,low}} = \begin{bmatrix} 1 & 0.05 & 0.05 \\ 0.05 & 1 & 0.05 \\ 0.05 & 0.05 & 1 \end{bmatrix}.$$

High covariance set-up,

$$\Sigma_{\text{placebo,high}} = \Sigma_{\text{active,high}} = \begin{bmatrix} 1 & 0.75 & 0.75 \\ 0.75 & 1 & 0.75 \\ 0.75 & 0.75 & 1 \end{bmatrix}.$$

4.5.2 Results

Simulations in a longitudinal setting with monotone missingness in the active arm only and a moderately large treatment effect of $\Delta = 0.3$ ($\mu_{a,3} = 2.2$ as observed in the asthma trial), show excellent approximation of the information anchored variance by Rubin's variance estimator for up to 40–50% missing data at the final follow-up, in each reference setting (Figure 4.5). The same results are observed when there is no treatment effect or a very large treatment effect of $\Delta = 1$. (see Figures E.1 and E.2).

With a smaller sample size of $n = 100$ per arm or a larger sample size of $n=1000$ per arm we still see excellent information anchoring by Rubin's variance estimator for up to 40–50% missing data in the active arm only for each reference based setting, see Figures 4.6 and 4.7 with $\Delta = 0.3$.

With a smaller covariance between the measurements at each time point j and j^* of $\sigma_{j,j^*} = 0.05$, for $j = 1, \dots, J$ and $j^* = 1, \dots, J$ where $j^* \neq j$, Rubin's variance estimator still approximates the information anchored variance for up to 40–50% missing data in the active arm, as shown in Figure 4.8. With a higher covariance between the measurements at each time point j and j^* of $\sigma_{j,j^*} = 0.75$, the approximation is excellent for up to 60–70% missingness, see Figure 4.9.

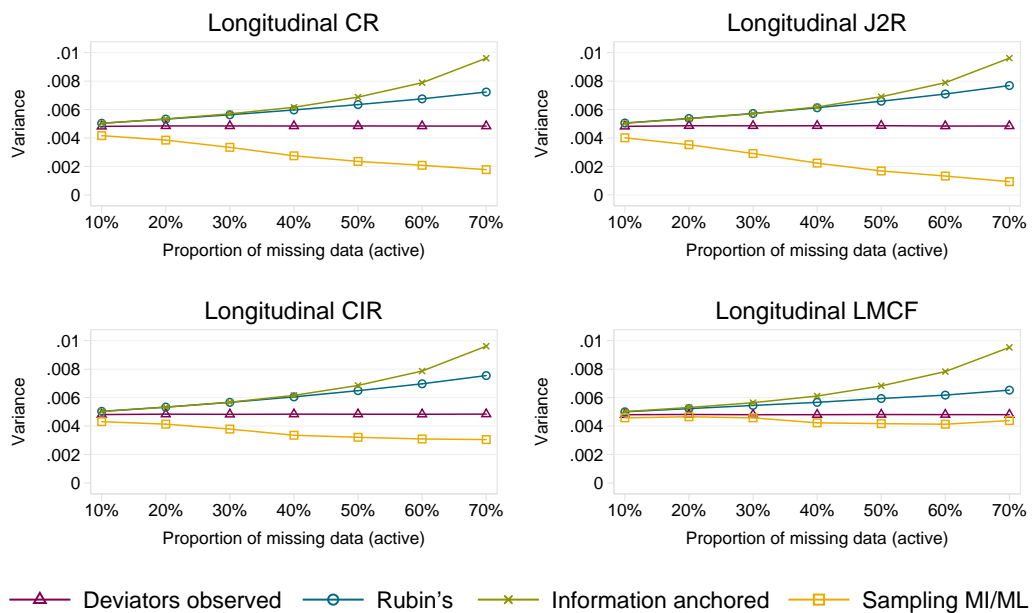


Figure 4.5: Rubin's variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with the asthma trial treatment effect of $\Delta = 0.3$ and $n=250$ per arm.

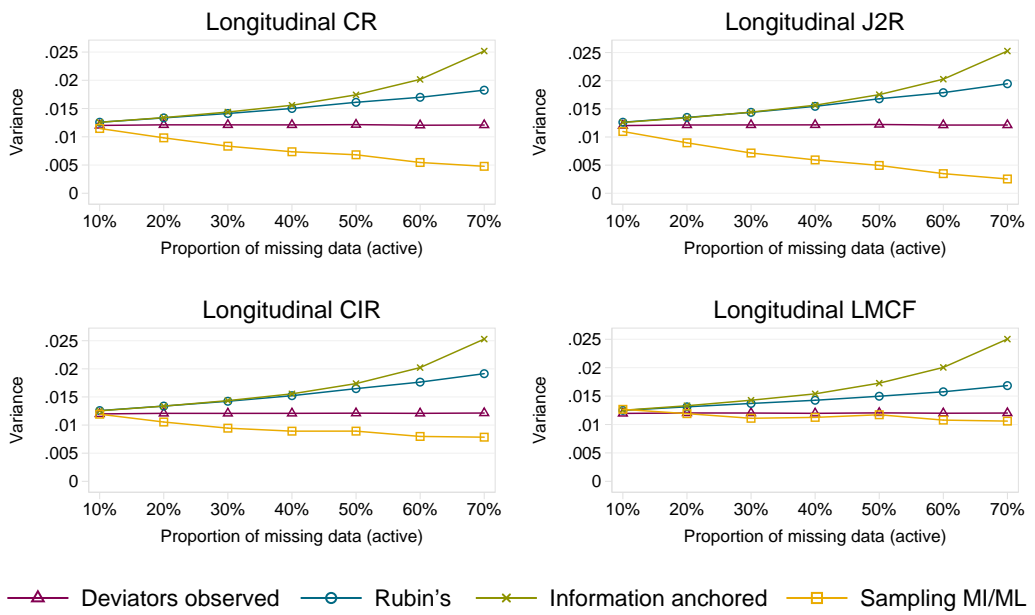


Figure 4.6: Rubin's variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with the asthma trial treatment effect of $\Delta = 0.3$ and $n=100$ per arm.

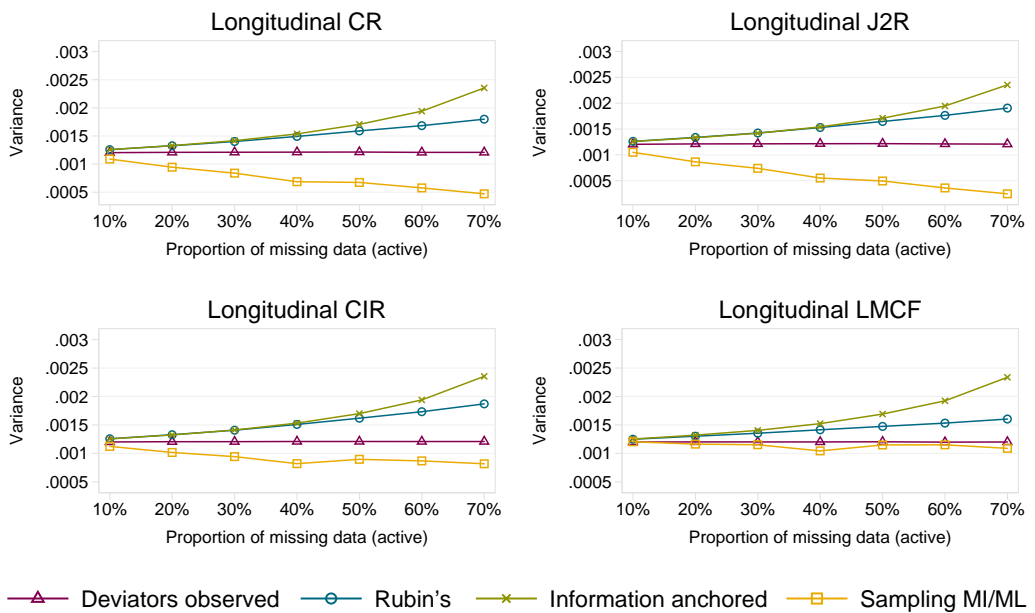


Figure 4.7: Rubin's variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with the asthma trial treatment effect of $\Delta = 0.3$ and $n=1000$ per arm.

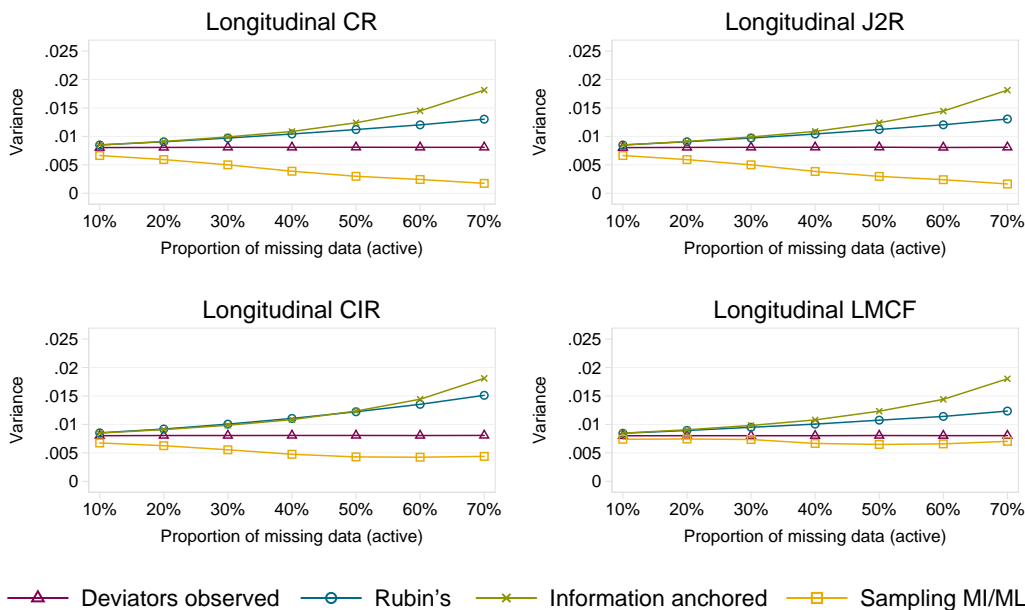


Figure 4.8: Rubin’s variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with the asthma trial treatment effect of $\Delta = 0.3$, $n=250$ per arm and a low covariance structure ($\sigma_{j,j^*} = 0.05$).

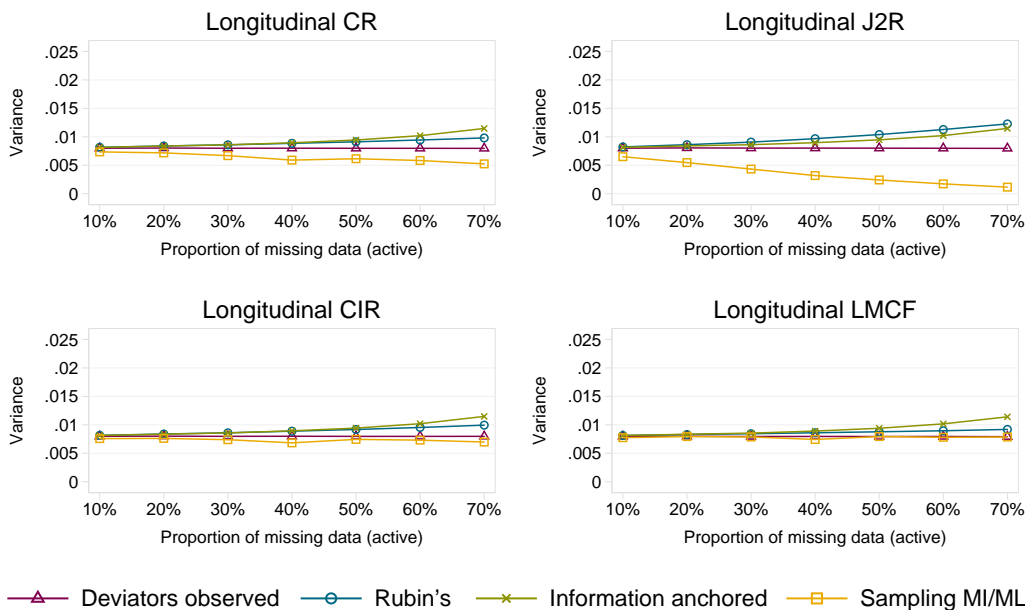


Figure 4.9: Rubin’s variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with the asthma trial treatment effect of $\Delta = 0.3$, $n=250$ per arm a high covariance structure ($\sigma_{j,j^*} = 0.75$).

When deviations occur at a similar rate in both arms we see Rubin’s variance estimator still provides an excellent approximation of the information anchored variance for realistic amounts of missingness. Figure 4.10 shows excellent agreement for up to 30–40% missing data in both arms. Beyond 40% missing data in both arms, the approximation is not so close. However we

note that this total amount of missingness would be uncommon in real trials. With a smaller or larger sample size per arm or a higher or lower correlation structure the approximation is still excellent for up to 30–40% missing data in both arms (see Appendix E.2 Figures E.5 to E.3). The results illustrate how with deviation in both arms, the approximation is additionally affected by the covariance between the parameters of the active and reference arm imputation models, as well as the proportion of missing data.

When half the proportion of deviation in the active arm is observed in the reference arm (Figure 4.11) we see the information anchoring by Rubin’s variance estimator is improved for missingness amounts above 40% in comparison to the setting with the same amount of deviation in both arms (Figure 4.10). Due to the smaller $\pi_{d,r}$, the terms in the difference between Rubin’s variance estimator and the information anchored variance (4.17) are smaller than with the same deviation in both arms.

We also looked at how well Rubin’s variance estimator approximates the information anchored variance for the baseline adjusted treatment effect in each of the longitudinal settings outlined above. In all cases the baseline adjustment had no notable impact on the information anchoring properties of Rubin’s variance estimator. There was excellent preservation of the information loss in the primary analysis in sensitivity analysis by Rubin’s variance estimator for realistic proportions of missing data in each reference based setting (results not shown).

Further simulations were run in a longitudinal setting, inspired by the asthma trial, with baseline and three follow-up time points, all subject to non-response, where the treatment effect of interest was the mean treatment group difference at the last follow-up time. There was again excellent preservation of the information loss in the primary analysis across sensitivity analysis by Rubin’s variance estimator for up to 40–50% missing data in the active arm only at the final follow-up and up to 30–40% missing data in both treatment arms, in each reference based setting as expected (results not shown for similarity).

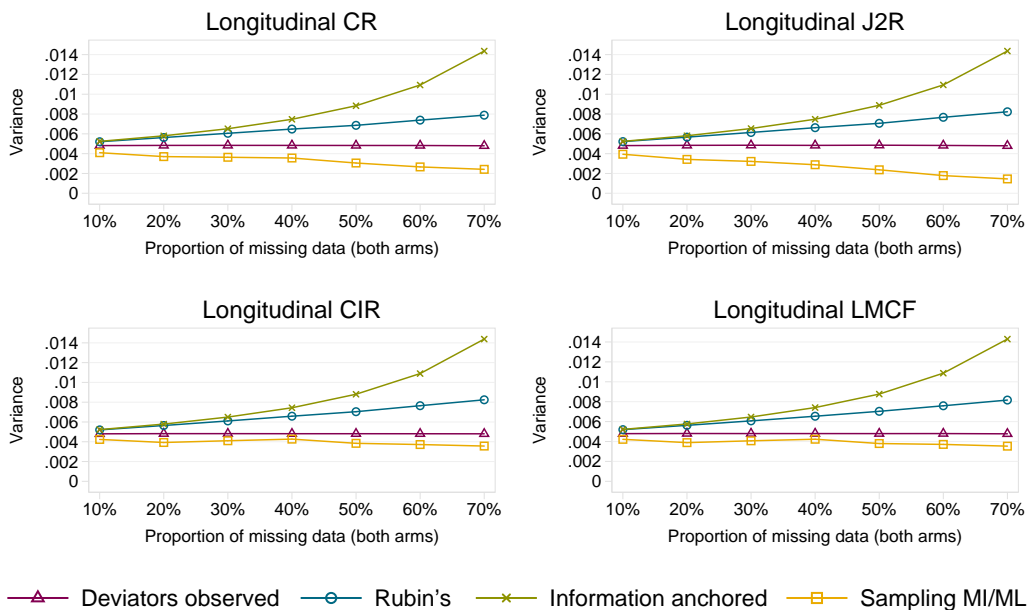


Figure 4.10: Rubin's variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with the asthma trial treatment effect of $\Delta = 0.3$, $n=250$ per arm and deviation in both treatment arms.

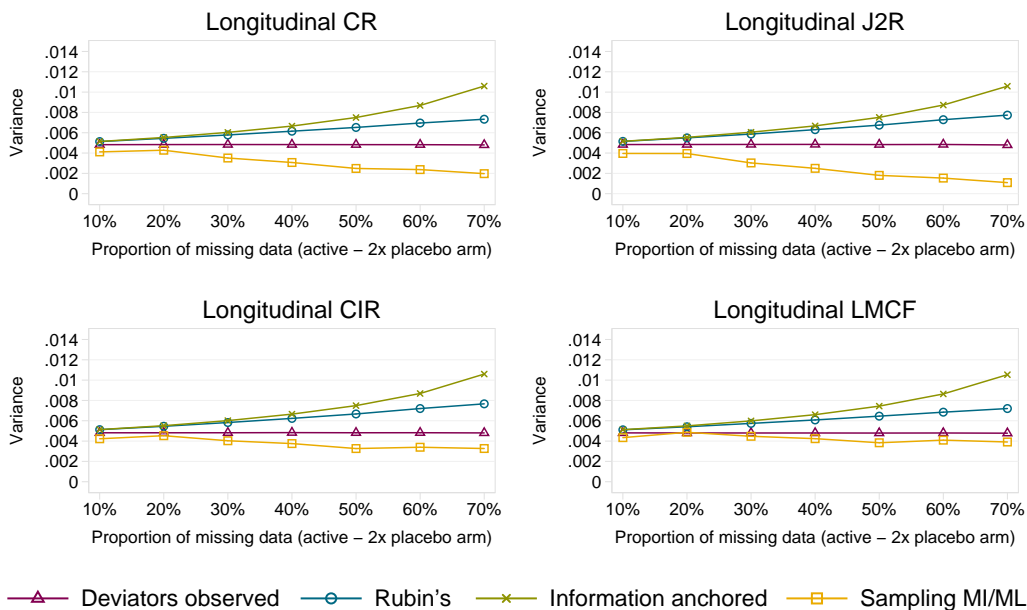


Figure 4.11: Rubin's variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with the asthma trial treatment effect of $\Delta = 0.3$, $n=250$ per arm and deviation active = 2x deviation placebo.

4.5.3 Discussion

We have looked at various longitudinal trial settings with different effect sizes, covariance structures and monotone missingness patterns via simulation. Rubin’s variance estimator is excellent at information anchoring in all realistic settings. The results correspond with the inferences drawn for the baseline and single follow-up setting. The additional complexity of an extra follow-up with non-response does not greatly impact the information anchoring performance of Rubin’s variance estimator in practice.

The results of the simulations are not surprising, given general Proposition 3. They are in line with the theory. In longitudinal settings with monotone missingness Rubin’s variance approximates the information anchored variance which preserves the loss of information in the primary analysis in reference based sensitivity analysis. The simulation results support the theoretical findings.

The proportion of missingness in each arm has a notable impact on the sharpness of the approximation. With deviation in one arm only it is only with very high amounts of missingness at the final follow-up (greater than 50%) when the approximation is not so sharp. However such high amounts of missingness in one arm are unexpected in practice. With deviation in both arms only it is only with very high amounts of missingness in both arms at the final follow-up (greater than 40%) when the approximation is not so sharp. With deviation in both arms the strength of the approximation is additionally effected by the covariance between the parameters in the active and reference imputation models.

4.6 Summary

The theory in this chapter shows that for reference based sensitivity analyses which satisfy $V_{\text{DF, full}} = V_{\text{DJ, full}} + O(n^{-2})$, estimation and inference via MI using the general algorithm given by Carpenter, Roger and Kenward [13] keeps the information approximately constant across the primary and sensitivity analysis. That is Rubin’s variance estimator approximately preserves the loss of information seen in the primary analysis in the sensitivity analysis. Simulations confirmed this is the case.

We have developed a general theory which defines the approximation. The difference in the variance of the parameters in the de-jure and de-facto imputation models, the covariance between these parameters and the proportion of missing data were identified as key factors which affect the sharpness of the approximation. We focussed our attention on either the unadjusted treatment group difference or the baseline adjusted treatment difference at the final follow-up. However, there may be many settings which satisfy $V_{\text{DF, full}} = V_{\text{DJ, full}} + O(n^{-2})$, thus for which the results hold.

Simulations revealed the approximation is not so sharp when the missing data pattern varies considerably by arm, i.e. with more than 50% missing data at the final follow-up in the active arm only. The strength of the approximation is affected by the amount of missingness and the discrepancy between the variance of the de-jure and de-facto imputation parameters. However we note that more than 50% missing data in a single trial arm generally would be rare in the clinical trial arena.

When there is deviation in both arms of a two arm trial, the approximation is not so sharp with more than 40% deviation in each arm. The strength of the approximation is affected by the amount of missingness and the discrepancy between the variance of the de-jure and de-facto imputation parameters for both the active and reference arm and further the covariance between the imputation parameters used in each arm. However we also note that in practice it is uncommon to see such high amounts of missing data in both arms. Thus trialists can be confident that Rubin's variance estimator provides a valid estimate of variance when using reference based sensitivity analysis.

Throughout this chapter and the previous we have followed the reference based MI algorithm described by Carpenter, Roger and Kenward [13], where the pre-deviation data, including baseline are modelled separately in each arm. Carpenter, Roger and Kenward acknowledge that "for appropriate post-deviation profiles, we therefore need the baseline means to be clinically similar (ideally equal) across arms. Randomization should ensure this." Where this is a concern they recommend centering baseline at the sample mean or sharing the baseline-time interaction across treatment arms. We remark that if the second approach is taken the presented results will still hold.

There will naturally be covariance between some of the imputation parameters of the active and reference de-jure imputation models and a greater covariance between the imputation parameters of the active and reference de-facto imputation models when the baseline-time interaction is shared. Thus in the most general longitudinal setting we will see slightly increased covariance terms in (4.17). Rubin's variance estimator will however still be to a good approximation information anchoring. We conclude we can use Rubin's variance estimator with confidence for information anchored sensitivity analysis.

Chapter 5

Sensitivity analysis using the ‘ δ -method’

The ‘ δ -method’ is an alternative controlled MI approach for sensitivity analysis of clinical trials with missing data. We continue with the setting of previous chapters; continuous baseline and outcome measure which can be modelled using the MVN distribution. We begin this chapter by describing the ‘ δ -method’ in full for sensitivity analysis of a longitudinal continuous outcome in Section 5.1. We then present our motivation for investigating Rubin’s treatment and variance estimator in this setting in Section 5.2. We apply our information anchoring principle for variance estimation in design based sensitivity analysis that was introduced in Section 3.1.

In Section 5.3 we then proceed to analytically investigate Rubin’s MI variance estimator for a trial with baseline and a single follow-up and compare this with our target information anchored variance. We focus on the setting with deviation in one arm only. Subsequently, we present a general proposition on the information anchoring properties of Rubin’s variance estimator in this context in Section 5.4. In Section 5.5 a supporting simulation study is presented. We then provide a more general proposition for Rubin’s δ -adjusted variance estimator in longitudinal trial settings in Section 5.6. A second simulation study is presented in Section 5.7 to support the theoretical results. In Section 5.8 we extend the results for deviation in both arms. We finish with a summary of the findings in Section 5.9.

5.1 The ‘ δ -method’

The ‘ δ -method’ allows trialists to assess the impact of deviators having a poorer/better response post-deviation than what is predicted for them under the MAR assumption, on the results of the primary analysis. That is a poorer/better response than those observed in the trial with the same history. The primary analysis is conducted under the MAR assumption, where the conditional distributions of later response data given earlier response data are the same for patients who do and do not deviate. The sensitivity analysis builds around the primary MAR model. The parameters of the conditional distributions used in imputation are constructed using the parameters from

the implied MAR conditional model, combined with our sensitivity analysis parameter, which is defined as δ .

Specifically, for sensitivity analysis, the conditional distributions for post-deviation data implied by MAR are edited by adding δ to the imputed post-deviation value. Suppose we have a trial with a single follow-up, all imputed post-deviation values may be edited by adding δ . Alternatively we may have a different δ for each treatment arm. For treatment arm z we can define δ_z as the change in response post-deviation (departure from MAR) in arm z and edit the imputed values accordingly. In a trial with two or more treatment arms the ‘ δ -method’ can be used to explore differences in post-deviation response by arm. Clearly, if desired, we can also explore the effect of a poorer response post-deviation in an experimental arm only.

For longitudinal trials we can add δ_z to the first post-deviation imputed value, $2\delta_z$ to the second and so on. This approach, outlined by Carpenter and Kenward [7, 44], is illustrated in Figure 1.1. δ_z is defined to represent the change in the rate of response post-deviation as compared to MAR, in treatment arm z . Or we can adjust imputations at all time points in arm z by δ_z . Alternatively a different δ may be specified for each missing data pattern. This of course requires the specification of many parameters.

Expert opinion and subject matter knowledge guides the choice of the sensitivity analysis parameters δ_z . They may be fixed or have their own prior with specified mean and variance and thus vary over the imputation set K , i.e. $\delta_{z,k} \sim N(\delta_z, \sigma_{\delta_z}^2)$ for treatment arm z and imputation k . White et al. [28] demonstrate successful elicitation of prior information on the difference between missing and observed outcomes using a number of experts and a pre-specified questionnaire. If the δ_z ’s are given a distribution, the covariance of the δ_z ’s must also be specified. It can be hard to elicit reliable information about this. A complicating factor is that the standard error of the final treatment estimate depends critically on the specified covariance of the δ_z ’s. Carpenter and Kenward [7] show that larger standard errors are obtained when the correlation is zero. This can be a useful starting point.

A different implementation of the δ -adjusted approach to sensitivity analysis progressively increases δ_z from 0 until the conclusions from the primary analysis are overturned. Each increment represents an increased departure from MAR ($\delta_z = 0$). If the value of δ_z that changes the conclusions is implausible i.e. realistically the missing data will not be that different from the observed data then greater confidence in the primary results can be inferred. This is referred to as a tipping point analysis by Yan [30]. In trials, the plausibility of specific values of δ_z should be assessed a-priori.

The general ‘ δ -method’ thus follows the pattern mixture modelling route. Following MI from the appropriately formed conditional distributions the intervention effect is re-estimated for each imputed data set and results combined using Rubin’s rules [19].

Recall that for imputation under MAR, for each deviating patient i , we require the conditional distribution of their missing data given their observed data, denoted as $[\mathbf{Y}_{Mi}|\mathbf{Y}_{Oi}, \boldsymbol{\eta}]$. $\boldsymbol{\eta}$ represents the parameters of this distribution. We obtain a unique draw of $\boldsymbol{\eta}$ from the appropriate posterior for each imputed data set required, before drawing missing data in turn. For MNAR imputation, for each deviating patient with missing data pattern m_i we require the distribution of their missing outcomes, given their observed data denoted as, $[\mathbf{Y}_{Mi}|\mathbf{Y}_{Oi}, m_i, \boldsymbol{\eta}_m]$ where $\boldsymbol{\eta}_m$ represents the the parameters of this distribution. For a continuous outcome the generic ‘ δ -method’ can then be summarised as follows,

1. Take all the observed data, and assuming MAR, fit a MVN distribution by treatment arm to obtain an estimate of the parameters of the conditional distributions of the missing data given the observed data, $\hat{\boldsymbol{\eta}}$.
2. For imputation k , draw $\boldsymbol{\eta}$ and δ_z , for $z = 1, \dots, N$ treatment arms, from their appropriate Bayesian posteriors.
3. Use the draws in step 2 to form $\boldsymbol{\eta}_m$, for each missing data pattern m , by adding or subtracting the draw of δ_z as required from the current draw of $\boldsymbol{\eta}$.
4. Impute the missing data using the current draw of $\boldsymbol{\eta}_m$, (from $[\mathbf{Y}_{Mi}|\mathbf{Y}_{Oi}, m_i, \boldsymbol{\eta}_m]$).
5. Repeat steps 2-4 K times.

Equivalently, and comparatively more simply, we could add the δ_z after imputation under MAR. The imputed datasets are each analysed using the substantive analysis model of interest (used in the primary analysis), and results combined using Rubin’s rules for inference. As the analysis model used in the primary analysis is retained in the sensitivity analysis the ‘ δ -method’ follows the *design based analysis* approach.

We note that other variants of the ‘ δ -method’ exist [60]. The developments explored here are directly relevant to other versions of the ‘ δ method’, which we do not consider further here.

The outlined ‘ δ -method’ evidently shares some similarities with reference based sensitivity analysis. The parameters from an observed data model under MAR form the essential building blocks in the imputation and a design based analysis approach is followed. However quantitative assumptions regarding the departures from MAR are made, rather than qualitative assumptions entailing the ‘borrowing’ of information between groups in the trial.

5.2 Motivation for variance evaluation

The ‘ δ -method’ uses an imputation model that makes different assumptions to the analysis model. For example, the imputation model may assume deviators have worse outcomes than those observed with the same history. The design based analysis model will not account for this variation in behaviour. As alluded to in Section 1.4 the frequency validity of Rubin’s rules long-run is not guaranteed when the imputation and substantive model have different structure, i.e. they are uncongenial. It is therefore also important to assess what Rubin’s rules are providing us with in this setting.

Since we are not borrowing information from one trial arm to impute another, the sampling variance of the δ -adjusted treatment effect does not exhibit the strange behaviour we observe in the reference based settings. In the δ -adjusted setting there is no imposed covariance between patients of opposing treatment arms. We therefore do not end up incorporating any covariance between patients of opposing treatment arms in the sampling variance of the δ -adjusted treatment effect. This appears in the reference based settings as shown in Subsection 2.2.2.

In Chapters 2 and 3 we introduced principles for variance estimation in sensitivity analysis where the design based analysis model used in the primary analysis is retained in the sensitivity analysis. That is, the estimated variance if we observed all the post-deviation data under the specified scenario provides a lower bound for variance estimation in sensitivity analysis. Here, once again, we desire to keep the information loss due to missing data constant or *anchored* across primary and all sensitivity analysis. This is a natural principle which respects the loss of information in the primary design based analysis and accordingly accounts for the uncertainty due to the information loss in the sensitivity analysis.

Throughout this chapter we retain the notation introduced in earlier chapters where $\hat{\theta}$ denotes the estimated treatment effect of interest from the primary design based analysis and the estimated variance of the treatment effect is denoted by V . Estimates indexed with the subscript $_{DJ}$ denotes analysis under our de-jure on-treatment MAR assumption. Estimates with the subscript $_{\delta}$ denote a δ -adjusted estimate.

Let $V_{_{DJ}, \text{full}}$ denote the variance estimator for the treatment effect from the design based analysis with full data i.e. no deviations; when deviations occur and post-deviation data are unobserved let $V_{_{DJ}, \text{MI}}$ denote Rubin's variance estimator for the treatment effect from the design based MAR MI analysis (primary analysis) and let $V_{_{\delta}, \text{full}}$ denote the variance estimator for the treatment effect from the design based analysis if post-deviation data incorporating a change in rate of response post-deviation (i.e. δ -adjusted) were actually observed. The required *information anchored* variance estimator for the treatment estimate in δ -adjusted sensitivity analysis denoted by V_{anchored} is,

$$V_{\text{anchored}} = \frac{V_{_{DJ}, \text{MI}}}{V_{_{DJ}, \text{full}}} \times V_{_{\delta}, \text{full}}.$$

We investigate whether Rubin's MI variance estimator in δ -adjusted sensitivity analysis, denoted by $V_{_{\delta}, \text{MI}}$, provides an acceptable estimate of the information anchored variance that preserves the loss of information in the primary analysis.

In the following section we derive the exact properties of Rubin's treatment and estimator and variance estimator in the baseline and single follow-up setting with δ -adjustment. We then compare Rubin's variance estimator to the information anchored variance.

5.3 Variance estimation in the baseline and single follow-up setting

In order to better understand what Rubin's rules are estimating for the treatment effect obtained using the ' δ -method', Rubin's MI estimate and variance estimator are now examined in a simple trial setting with baseline and a single follow-up. This is the same setting considered in Chapter 2, when the reference based treatment effect was examined. Throughout, we assume the data is normal and the variance is known. As discussed in Subsection 3.1.1, following inference by Carpenter and Kenward [44] this does not have a material impact on the conclusions and is asymptotically irrelevant.

5.3.1 Rubin's variance estimator under the 'δ-method'

As in Chapter 2, consider a two arm randomised trial, with treatment group denoted by $z = a, r$ with $i = 1, \dots, n$ subjects per arm, with a baseline (time 1) and single follow-up outcome (time 2). The design based analysis is the difference in means between the two trial arms at time 2. All of the reference arm and n_o of the n active arm are observed at both times without deviation. Post-baseline the remaining n_d active arm patients deviate such that in the active arm $n_o + n_d = n$. We condition on n_d and where \mathcal{O} and \mathcal{D} define the sets of indices for the completers and deviators in the active arm, assume the data comes from the following bivariate normal distributions with known variance,

$$\begin{pmatrix} Y_{zi1} \\ Y_{zi2} \end{pmatrix} \sim N \left\{ \begin{pmatrix} \mu_{z1} \\ \mu_{z2} \end{pmatrix}; \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \right\}, \quad z = a, r, \quad i = 1, \dots, n.$$

In Subsection 3.1.1 the expectation of Rubin's MI variance estimator for the treatment estimate, under the MAR assumption, was derived in this setting. To derive the expectation of Rubin's MI variance estimator for the treatment estimate using the 'δ-method' we follow a similar approach, employing separate imputation models by randomised arm. An appropriate imputation distribution for the missing outcomes at time 2 is required, with a suitable posterior for the included parameters. To simplify subsequent calculations we define $Y_{ai1}^* = Y_{ai1} - \bar{Y}_{a1,o}$ where $\bar{Y}_{a1,o} = \frac{1}{n_o} \sum_{i \in \mathcal{O}} Y_{ai1}$.

Under MAR, our imputation model is formed from the regression of Y_{ai2} on Y_{ai1}^* for the set of completers in the active arm ($i \in \mathcal{O}$). Subsection 3.1.1 summarised the ordinary least squares estimates for this fitted model. We assume the data is normal, the variance is known and improper priors for the missing observations themselves, so the posterior of the model parameters is approximately normal and let the observed data dominate the prior. Thus, as presented in Subsection 3.1.1, the MAR imputation model for patient i and imputation k is,

$$\tilde{Y}_{ai2,k} = \bar{Y}_{a2,o} + u_k + (r/q + b_k) (Y_{ai1} - \bar{Y}_{a1,o}) + e_{i,k}, \quad \text{for } i \in \mathcal{D}, k = 1, \dots, K,$$

where,

$$\begin{aligned} u_k &\sim N(0, n_o^{-1} \sigma_{2.1}) \\ b_k &\sim N(0, q^{-1} \sigma_{2.1}) \\ e_{i,k} &\sim N(0, \sigma_{2.1}). \end{aligned}$$

$\bar{Y}_{a2,o} = \frac{1}{n_o} \sum_{i \in \mathcal{O}} Y_{ai2}$, $r = \sum_{i \in \mathcal{O}} (Y_{ai1} - \bar{Y}_{a1,o}) (Y_{ai2} - \bar{Y}_{a2,o})$ and $q = \sum_{i \in \mathcal{O}} (Y_{ai1} - \bar{Y}_{a1,o})^2$. For imputation using the 'δ-method' we must now add δ_k to the model, where δ_k represents the change in response post-deviation for imputation k . With the additional assumption that $\delta_k \sim N(\delta, \sigma_\delta^2)$

we have the required imputation model as,

$$\tilde{Y}_{ai2,k} = \bar{Y}_{a2o} + \delta_k + u_k + (r/q + b_k) (Y_{ai1} - \bar{Y}_{a1o}) + e_{i,k}, \text{ for } i \in \mathcal{D}, k = 1, \dots, K.$$

We are interested in the expectation of Rubin's MI estimator and variance estimator for the average treatment group difference at time 2, which we denote respectively as $\hat{\theta}_{\delta, \text{MI}}$ and $V_{\delta, \text{MI}}$. Firstly,

$$E(\hat{\theta}_{\delta, \text{MI}}) = E\left(\frac{1}{K} \sum_{k=1}^K \hat{\theta}_{\delta,k}\right).$$

From the imputation model, we obtain the sample mean treatment difference, at time 2 in the k^{th} imputation set, $\hat{\theta}_{\delta,k}$ as,

$$\hat{\theta}_{\delta,k} = \hat{\mu}_{a2,k} - \hat{\mu}_{r2} = \frac{n_o}{n} \bar{Y}_{a2o} + \frac{n_d}{n} \bar{Y}_{a2d,k} - \bar{Y}_{r2},$$

where,

$$\bar{Y}_{a2d,k} = \frac{1}{n_d} \sum_{i \in \mathcal{D}} \tilde{Y}_{ai2,k} = \bar{Y}_{a2,o} + \delta_k + u_k + \left(\frac{r}{q} + b_k\right) (\bar{Y}_{a1d} - \bar{Y}_{a1,o}) + \bar{e}_k,$$

and $\bar{e}_k = \frac{1}{n_d} \sum_{i \in \mathcal{D}} e_{i,k}$ and $\bar{Y}_{a1d} = \frac{1}{n_d} \sum_{i \in \mathcal{D}} Y_{ai1}$. Averaging the sample mean estimate for the k^{th} imputation, over the imputation set, K , gives Rubin's MI point estimator as,

$$\hat{\theta}_{\delta, \text{MI}} = \frac{n_o}{n} \bar{Y}_{a2o} + \frac{n_d}{n} \left[\bar{Y}_{a2,o} + \bar{\delta} + \bar{u} + \left(\frac{r}{q} + \bar{b}\right) (\bar{Y}_{a1d} - \bar{Y}_{a1,o}) + \bar{e} \right] - \bar{Y}_{r2},$$

where $\bar{\delta} = \frac{1}{K} \sum_{k=1}^K \delta_k$, $\bar{u} = \frac{1}{K} \sum_{k=1}^K u_k$, $\bar{b} = \frac{1}{K} \sum_{k=1}^K b_k$ and $\bar{e} = \frac{1}{K} \sum_{k=1}^K \bar{e}_k$. Taking the necessary expectations reveals,

$$E(\hat{\theta}_{\delta, \text{MI}}) = \frac{n_o}{n} \mu_{a2} + \frac{n_d}{n} (\mu_{a2} + \delta) - \mu_{r2} = \mu_{a2} + \frac{n_d}{n} \delta - \mu_{r2}.$$

Rubin's treatment estimator is unbiased for the δ -adjusted treatment estimate. This is unsurprising given the results of the previous chapters. The imputation incorporates δ_k and additive terms which have an expectation of zero, thus taking the average treatment effect over the imputation set provides an unbiased estimator. Going forward we therefore focus our evaluations on the variance estimator. For Rubin's MI variance estimator we require,

$$E(V_{\delta, \text{MI}}) = E(\hat{W}) + \left(1 + \frac{1}{K}\right) E(\hat{B}),$$

where, \hat{W} and, \hat{B} are defined as in (1.7) and (1.8). First we consider $E(\hat{W})$,

$$\begin{aligned} E(\hat{W}) = E(\hat{\sigma}_k^2) = & E\left(\frac{\frac{1}{n-1} \sum_{i=1}^n (Y_{ri2} - \hat{\mu}_{r2})^2}{n}\right) \\ & + E\left(\frac{\frac{1}{n-1} \left[\sum_{i \in \mathcal{O}} (Y_{ai2} - \hat{\mu}_{a2,k})^2 + \sum_{i \in \mathcal{D}} (\hat{Y}_{ai2,k} - \hat{\mu}_{a2,k})^2\right]}{n}\right). \end{aligned} \quad (5.1)$$

We define the first component on the RHS of (5.1) as $W1$ and the second component on the RHS of (5.1) as $W2$. Expanding these sums where required and taking the necessary expectations, (see Appendix F.2 for detailed calculations) for time 1 a baseline, where $\mu_{a1} = \mu_{r1}$, $\pi_d = n_d/n$ and $\rho^2 = \sigma_{12}^2/(\sigma_{11}\sigma_{22})$ (the correlation between time 1 and time 2 squared) gives,

$$E(W1) = \frac{\sigma_{22}}{n},$$

$$E(W2) = \frac{\sigma_{22}}{n} + \frac{2n_d\sigma_{22}(1-\rho^2)}{n(n_o-1)(n-1)} + \frac{n_o\pi_d(\delta^2 + \sigma_\delta^2)}{n(n-1)}. \quad (5.2)$$

We now consider $(1 + \frac{1}{K}) E(\hat{B})$. Taking the necessary expectations reveals,

$$\left(1 + \frac{1}{K}\right) E(\hat{B}) = \left(1 + \frac{1}{K}\right) \left(\frac{\sigma_{22}(1-\rho^2)n_d}{n(n_o-1)} + \pi_d^2\sigma_\delta^2\right). \quad (5.3)$$

Assuming that n is sufficiently large so that we may take $(n-1)$ to be n and (n_o-1) to be n_o this gives Rubin's variance estimator asymptotically as,

$$E[V_{\delta, \text{MI}}] = \frac{2\sigma_{22}}{n} + \frac{2\sigma_{22}(1-\rho^2)\pi_d}{n^2(1-\pi_d)} + \frac{\pi_d(1-\pi_d)(\delta^2 + \sigma_\delta^2)}{n} + \left(1 + \frac{1}{K}\right) \left(\frac{\sigma_{22}(1-\rho^2)\pi_d}{n(1-\pi_d)} + \pi_d^2\sigma_\delta^2\right).$$

For infinite K this is equivalent to,

$$E[V_{\delta, \text{MI}}] = \sigma_{22} \left[\frac{1}{n} + \frac{1}{n_o} \right] - \frac{\sigma_{12}^2}{\sigma_{11}} \frac{\pi_d}{n(1-\pi_d)} + \frac{2\sigma_{22}(1-\rho^2)\pi_d}{n^2(1-\pi_d)} + \frac{\pi_d(1-\pi_d)(\delta^2 + \sigma_\delta^2)}{n} + \pi_d^2\sigma_\delta^2. \quad (5.4)$$

Table F.1 summarises Rubin's derived variance with δ -adjustment alongside simulation results. Derived results correspond with the simulation results (within 2 MCSE). Under the fixed δ assumption, that is under the assumption that $\delta_k \sim N(\delta, 0)$ for Rubin's MI variance estimator we have,

$$E(W1) = \frac{\sigma_{22}}{n} + \frac{2n_d\sigma_{22}(1-\rho^2)}{n(n_o-1)(n-1)} + \frac{n_o\pi_d\delta^2}{n(n-1)}.$$

For the between imputation variance,

$$\left(1 + \frac{1}{K}\right) E(\hat{B}) = \left(1 + \frac{1}{K}\right) \frac{\sigma_{22}(1-\rho^2)n_d}{n(n_o-1)}.$$

Assuming that n is sufficiently large so that we may take $(n-1)$ to be n and (n_o-1) to be n_o this gives Rubin's estimate asymptotically as,

$$E[V_{\delta, \text{MI}}] = \frac{2\sigma_{22}}{n} + \frac{2\sigma_{22}(1-\rho^2)\pi_d}{n^2(1-\pi_d)} + \frac{\pi_d(1-\pi_d)\delta^2}{n} + \left(1 + \frac{1}{K}\right) \frac{\sigma_{22}(1-\rho^2)\pi_d}{n(1-\pi_d)}.$$

For infinite K this is equivalent to,

$$E[V_{\delta, \text{MI}}] = \sigma_{22} \left[\frac{1}{n} + \frac{1}{n_o} \right] - \frac{\sigma_{12}^2}{\sigma_{11}} \frac{\pi_d}{n(1-\pi_d)} + \frac{2\sigma_{22}(1-\rho^2)\pi_d}{n^2(1-\pi_d)} + \frac{\pi_d(1-\pi_d)\delta^2}{n}.$$

Table F.2 summarises Rubin's derived variance with fixed δ adjustment alongside simulation re-

sults. Derived results again correspond with simulation results (within 2 MCSE).

5.3.2 Information anchoring

The required information anchored variance, that preserves the information loss seen in the primary analysis under MAR in δ -adjusted sensitivity analysis, is given by,

$$V_{\text{anchored}} = \frac{V_{\text{DJ, MI}}}{V_{\text{DJ, full}}} \times V_{\delta, \text{full}}.$$

As derived in Subsection 3.1.1, the expectation of the MI MAR variance for $K \rightarrow \infty$ is,

$$E[V_{\text{DJ, MI}}] = \sigma_{22} \left[\frac{1}{n} + \frac{1}{n_o} \right] - \frac{\sigma_{12}^2}{\sigma_{11}} \frac{\pi_d}{n(1-\pi_d)} + \frac{2\sigma_{22}(1-\rho^2)\pi_d}{n^2(1-\pi_d)}.$$

As in Section 3.2, to simplify subsequent calculations we take up to the $O(n^{-2})$ terms for $E[V_{\text{DJ, MI}}]$ since this is equivalent to the MAR ML variance. The expectation of $V_{\text{DJ, full}}$ using standard normal distribution results is,

$$E[V_{\text{DJ, full}}] = \frac{2\sigma_{22}}{n}.$$

In the current δ -adjusted setting the expected value of the design based variance estimator, were we able to observe the post-deviation data is (see Appendix F.1 for detailed calculations),

$$E[V_{\delta, \text{full}}] = \frac{2\sigma_{22}}{n} + \frac{n_o n_d \delta^2}{n^3}.$$

Therefore the expectation of the information anchored variance is,

$$E[V_{\text{anchored}}] = \frac{2\sigma_{22}}{n} + \frac{n_o n_d \delta^2}{n^3} + \frac{\delta^2(1-\rho^2)\pi_d^2}{2n} + \frac{\sigma_{22}(1-\rho^2)\pi_d}{n(1-\pi_d)}. \quad (5.5)$$

The difference between Rubin's MI variance estimator for the δ -adjusted (non-fixed) treatment effect and the information anchored variance, (5.4) - (5.5) is,

$$E[V_{\delta, \text{MI}}] - E[V_{\text{anchored}}] = \sigma_{\delta}^2 \pi_d^2 + \frac{2\sigma_{22}(1-\rho^2)\pi_d}{n^2(1-\pi_d)} - \frac{\delta^2(1-\rho^2)\pi_d^2}{2n}.$$

Which up to the order π_d^3/n^2 terms is,

$$E[V_{\delta, \text{MI}}] - E[V_{\text{anchored}}] = \sigma_{\delta}^2 \pi_d^2 + \frac{2\sigma_{22}(1-\rho^2)}{n^2} [\pi_d + \pi_d^2 + \pi_d^3] - \frac{\delta^2(1-\rho^2)\pi_d^2}{2n}.$$

We see the largest term in the difference between Rubin's MI variance estimator and the information anchored variance under δ -adjustment is, $\sigma_{\delta}^2 \pi_d^2 = \sigma_{\delta}^2 n_d^2/n^2$. We therefore see that the approximation is principally driven by the assumed variance for δ_k . Using Rubin's variance estimator, we will lose information relative to the information anchored variance if σ_{δ}^2 is large. The approximation is sharper for smaller σ_{δ}^2 . The other two terms are at least an order of n smaller. That is $O(\pi_d^2 n^{-1})$ or $O(n^{-2})$ and they form the difference when δ_k is fixed, i.e. $\delta_k \sim N(\delta, 0)$. To see this explicitly, when δ_k is fixed the two largest terms in the difference are,

$$\frac{2\sigma_{22}(1-\rho^2)\pi_d}{n^2} - \frac{\delta^2(1-\rho^2)\pi_d^2}{2n}.$$

Therefore the difference is of a smaller order when δ_k is fixed. In this case it is at most up to $O(\pi_d^2 n^{-1})$. The difference depends on the size of the trial (n), the amount of missingness in the active arm (n_d), the correlation between the time 1 and time 2 measurement (ρ^2), the variance of the data at time 2 (σ_{22}) and the specified size of δ . However, it is always of a very small magnitude. The information anchoring variance we are targeting is dominated by the $2\sigma_{22}/n$ term which is $O(n^{-1})$. These terms are tiny in comparison.

Thus with fixed δ the difference between Rubin's variance and the information variance is practically negligible. When δ is not fixed the information anchoring performance of Rubin's variance estimator is dependent on the proportion of missing data and the variance of δ_k . If there is a great deal of uncertainty surrounding the proposed difference between the deviators and those observed with the same history, i.e. large σ_{δ}^2 then the approximation will not be so sharp. Information anchoring will be better when δ_k has a small variance and there is a small proportion of missing data. In both cases (fixed or not fixed δ), Rubin's MI treatment estimate is unbiased.

5.4 General theory for Rubin's variance estimator; baseline and single follow-up

We now present a general theory on Rubin's variance estimator and how it anchors the information loss across the primary and sensitivity analysis in the baseline and single follow-up δ -adjusted setting. This consolidates the results of the derivation presented in Section 5.3. The more formal

theory also applies to the baseline adjusted treatment effect. We denote Rubin’s variance estimator for the treatment effect in the design based sensitivity analysis where post-deviation data are unobserved but imputed with δ -adjustment as $V_{\delta, \text{MI}}$. For ease of exposition the proposition focuses on the setting with deviation in one arm only. In Section 5.8 we show how the result holds with deviation in both arms, with and without a different δ -adjustment by trial arm.

5.4.1 Setting for Proposition 4

Consider a two arm trial consisting of $2n$ patients with a single continuous follow-up outcome and baseline. n patients have been randomised to an active arm and n to a reference arm. \mathbf{Y} is the $(2n \times 1)$ vector of continuous outcome data, and \mathbf{x} , is the corresponding $(2n \times p)$ matrix of observed baseline covariates, including baseline outcome and treatment. The estimated treatment effect can be written $\mathbf{a}^T \mathbf{Y}$ where \mathbf{a} may or may not incorporate \mathbf{x} .

We assume \mathbf{Y} is normally distributed and has known variance Σ . The variance of the data is assumed to be known without any serious loss of generality, following inference by Carpenter and Kenward in [44]. As discussed in Section 4.1, because of the conditional argument underlying the derivation this doesn’t have a profound impact on the exposition asymptotically.

Suppose all n reference patients and n_o of the active arm are observed without deviation at baseline and follow-up. The remaining n_d active arm patients deviate post-baseline such that $n_o + n_d = n$. Let \mathcal{D} and \mathcal{O} define the sets of indices for the deviators and completers in the active arm respectively. π_d denotes the proportion of deviators in the active arm, i.e. n_d/n .

Further $\delta_k \sim N(\delta, \sigma_\delta^2)$ and $V_{\delta, \text{full}} = V_{\mathcal{D}, \text{full}} + O(\pi_d(1 - \pi_d)n^{-1})$. Typically, although not necessarily, some form of regression model will be the substantive model of interest, used to make inferences about the treatment effect in the population of interest. Appendix F.1 shows equation (5.7) holds for the mean treatment difference, adjusted and unadjusted for baseline.

5.4.2 Proposition 4

For sensitivity analyses of a two arm trial with baseline and a single follow-up performed by MI using the ‘ δ -method’ as outlined in [44], the difference between Rubin’s variance estimator and the information anchored variance for the treatment effect is,

$$-\pi_d^2 \sigma_\delta^2 + \frac{E[\hat{B}_{\mathcal{D}, \text{J}}]}{E[\hat{W}_{\mathcal{D}, \text{J}}]} O\left(\frac{\pi_d(1 - \pi_d)}{n}\right).$$

In other words,

$$\frac{V_{\text{DJ, MI}}}{V_{\text{DJ, full}}} = \frac{V_{\delta, \text{MI}}}{V_{\delta, \text{full}}} - \pi_d^2 \sigma_\delta^2 + \frac{E[\hat{B}_{\text{DJ}}]}{E[\hat{W}_{\text{DJ}}]} O\left(\frac{\pi_d(1-\pi_d)}{n}\right).$$

When $\delta_k \sim N(\delta, 0)$, the difference between Rubin's variance estimator and the information anchored variance is,

$$\frac{E[\hat{B}_{\text{DJ}}]}{E[\hat{W}_{\text{DJ}}]} O\left(\frac{\pi_d(1-\pi_d)}{n}\right).$$

In other words,

$$\frac{V_{\text{DJ, MI}}}{V_{\text{DJ, full}}} = \frac{V_{\delta, \text{MI}}}{V_{\delta, \text{full}}} + \frac{E[\hat{B}_{\text{DJ}}]}{E[\hat{W}_{\text{DJ}}]} O\left(\frac{\pi_d(1-\pi_d)}{n}\right).$$

5.4.3 Proof of Proposition 4

Let $z = a, r$ index the randomised active arm or reference arm allocation for each patient i with follow-up outcome denoted by Y_{zi} . The outcome data for the reference patients are contained in the vector $\mathbf{Y}_r = (Y_{r1}, \dots, Y_{rn})^T$. The outcome data for the non-deviating active patients are contained in the vector $\mathbf{Y}_{a,o} = \{Y_{ai}; i \in \mathcal{O}\}^T$.

We suppose that each deviating patient has two potential outcomes: the one that would occur if they remain on active treatment (de-jure) and the other that would occur if they had a better/worse response than if they remained on active treatment, i.e. plus/minus some amount δ . The potentially observable de-jure data for the n_d deviating patients are contained in the vector $\mathbf{Y}_{a,DJ,d}$ and the alternative δ -adjusted outcome data in the vector $\mathbf{Y}_{a,\delta,d}$. Define the collection of observed and potentially observable outcome data as $\mathbf{Y} = (\mathbf{Y}_r, \mathbf{Y}_{a,o}, \mathbf{Y}_{a,DJ,d}, \mathbf{Y}_{a,\delta,d})^T$ which has dimensions $[(n + n_o + 2n_d) \times 1]$.

For each deviating patient we can only observe one of the potential outcomes, either de-jure or δ -adjusted. Consider two $[(n + n_o + 2n_d) \times (n + n_o + 2n_d)]$ matrices of 0's and 1's, \mathbf{D}_{DJ} and \mathbf{D}_δ , arranged such that $\mathbf{D}_{\text{DJ}}\mathbf{Y}$ gives the $[(n + n_o + 2n_d) \times 1]$ de-jure data (with zero entries for the potential δ adjusted outcomes). $\mathbf{D}_\delta\mathbf{Y}$ gives the $[(n + n_o + 2n_d) \times 1]$ δ -adjusted data (with zero entries for the potential de-jure, here MAR outcomes).

Let \mathbf{a} be a $[(n + n_o + 2n_d) \times 1]$ vector such that $\mathbf{a}^T \mathbf{D}_{\text{DJ}} \mathbf{Y}$ returns the de-jure treatment estimate and $\mathbf{a}^T \mathbf{D}_\delta \mathbf{Y}$ returns the δ -adjusted treatment estimate. When deviating patients experience de-jure behaviour post-deviation and their follow-up outcome data are observed the expectation of the variance estimator for the de-jure on-treatment estimand can be expressed as,

$$E[V_{D_J, \text{full}}] = E[V(\mathbf{a}^T \mathbf{D}_{D_J} \mathbf{Y})] = E[\mathbf{a}^T \mathbf{D}_{D_J} V(\mathbf{Y}) \mathbf{D}_{D_J}^T \mathbf{a}] = \mathbf{a}^T \mathbf{D}_{D_J} \Sigma \mathbf{D}_{D_J}^T \mathbf{a}. \quad (5.6)$$

We use the proposed condition that the expectation of the variance estimator for the δ -adjusted estimand can be expressed as,

$$E[V_{\delta, \text{full}}] = \mathbf{a}^T \mathbf{D}_{D_J} \Sigma \mathbf{D}_{D_J}^T \mathbf{a} + O\left(\frac{\pi_d(1 - \pi_d)}{n}\right). \quad (5.7)$$

We now suppose that post-deviation data are unobserved, i.e. the potentially observable de-facto and δ -adjusted entries in \mathbf{Y} are missing for the n_d active patients. We alternatively impute these outcomes, using de-jure imputation and δ imputation. This gives K ‘complete’ data samples \mathbf{Y}_k , of size $[(n + n_o + 2n_d) \times 1]$. For this we need an appropriate imputation distribution for the missing data under each scenario, with a suitable posterior for the included parameters.

Under our de-jure assumption (on treatment MAR) our imputation model is formed from the regression of $\mathbf{Y}_{a,o}$ on $\mathbf{P}_{a,o}$ where $\mathbf{P}_{a,o}$ is the $[n_o \times p]$ design matrix for the imputation model, which contains the values of the $(p-1)$ covariates included in the imputation model (including the baseline outcome but excluding the treatment indicator since we perform imputation separately by arm) with a vector of 1’s to include an intercept term in the model, for observed active patients.

The parameter estimates for the de-jure imputation model are found using, $\hat{\beta}_{a,o} = (\mathbf{P}_{a,o}^T \mathbf{P}_{a,o})^{-1} \mathbf{P}_{a,o}^T \mathbf{Y}_{a,o}$ with known covariance matrix $\mathbf{V}_{a,o} = (\mathbf{P}_{a,o}^T \mathbf{P}_{a,o})^{-1} \sigma^2$. We assume the large sample posterior for the parameter estimates of the de-jure imputation model, denoted as $\hat{\beta}_{D_J}$, is normal and centered on the ML estimator $\hat{\beta}_{a,o}$ with covariance matrix $\mathbf{V}_{a,o}$. That is,

$$\hat{\beta}_{D_J} | \mathbf{Y}_{a,o} \sim N(\hat{\beta}_{a,o}; \mathbf{V}_{a,o}).$$

The de-jure imputation model for patient i and imputation k can therefore be expressed as,

$$\tilde{Y}_{ai,k} | \mathbf{Y}_{a,o} = \mathbf{P}_{a,d,i} \left[\hat{\beta}_{a,o} + \mathbf{b}_{a,o,k} \right] + e_{i,k} \text{ for } i \in \mathcal{D},$$

where $\mathbf{b}_{a,o,k} \sim N(0, \mathbf{V}_{a,o})$, $e_{i,k} \sim N(0, \sigma^2)$ and $\mathbf{P}_{a,d,i}$ is the covariate data for each deviating patient i (excluding treatment group but including a 1), of dimensions $[1 \times p]$.

For δ -adjusted imputation we must add δ_k into the imputation model. The δ -adjusted imputation model for patient i and imputation k can therefore be expressed as,

$$\check{Y}_{ai,k} | \mathbf{Y}_{a,o} = \mathbf{P}_{a,d,i} \left[\hat{\beta}_{a,o} + \mathbf{b}_{a,o,k} \right] + \delta_k + e_{i,k} \text{ for } i \in \mathcal{D},$$

where $\mathbf{b}_{a,o,k} \sim N(0, \mathbf{V}_{a,o})$, $\delta_k \sim N(\delta, \sigma_\delta^2)$, and $e_{i,k} \sim N(0, \sigma^2)$. We are interested in imputation inference for, $\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_{\text{DJ}} \mathbf{Y}_k$ or $\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_\delta \mathbf{Y}_k$. Letting the number of imputations, $K \rightarrow \infty$, the variance of our MI treatment estimate as estimated by Rubin's rules is, $V_{\text{DJ, MI}} = \hat{W}_{\text{DJ}} + \hat{B}_{\text{DJ}}$ or $V_{\delta, \text{MI}} = \hat{W}_\delta + \hat{B}_\delta$ where,

$$E \left[\hat{W}_{\text{DJ}} \right] = E \left[\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_{\text{DJ}} \hat{\Sigma}_k \mathbf{D}_{\text{DJ}}^T \mathbf{a} \right] \rightarrow \mathbf{a}^T \mathbf{D}_{\text{DJ}} \Sigma \mathbf{D}_{\text{DJ}}^T \mathbf{a},$$

and,

$$E \left[\hat{W}_\delta \right] = E \left[\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_\delta \hat{\Sigma}_k \mathbf{D}_\delta^T \mathbf{a} \right] \rightarrow \mathbf{a}^T \mathbf{D}_{\text{DJ}} \Sigma \mathbf{D}_{\text{DJ}}^T \mathbf{a} + O \left(\frac{\pi_d(1 - \pi_d)}{n} \right).$$

Under de-jure,

$$\hat{B}_{\text{DJ}} = \frac{1}{K-1} \sum_{k=1}^K \left[\pi_d (\bar{e}_k - \bar{e}) + \pi_d (\bar{\mathbf{P}}_{a,d} \mathbf{b}_{a,o,k} - \bar{\mathbf{P}}_{a,d} \bar{\mathbf{b}}_{a,o}) \right]^2,$$

where $\bar{e}_k = \frac{1}{n_d} \sum_{i \in \mathcal{D}} e_{i,k}$, $\bar{e} = \frac{1}{K} \sum_{k=1}^K \bar{e}_k$, $\bar{\mathbf{P}}_{a,d} = \frac{1}{n_d} \sum_{i \in \mathcal{D}} \mathbf{P}_{a,d,i}$ and $\bar{\mathbf{b}}_{a,o} = \frac{1}{K} \sum_{k=1}^K \mathbf{b}_{a,o,k}$. This has expectation,

$$E \left[\hat{B}_{\text{DJ}} \right] = \pi_d^2 \left[\frac{\sigma^2 + n_d \bar{\mathbf{P}}_{a,d} \mathbf{V}_{a,o} \bar{\mathbf{P}}_{a,d}^T}{n_d} \right].$$

With δ -adjustment,

$$\hat{B}_\delta = \frac{1}{K-1} \sum_{k=1}^K \left[\pi_d (\bar{e}_k - \bar{e}) + \pi_d (\bar{\mathbf{P}}_{a,d} \mathbf{b}_{a,o,k} - \bar{\mathbf{P}}_{a,d} \bar{\mathbf{b}}_{a,o}) + \pi_d (\delta_k - \bar{\delta}) \right]^2,$$

where $\bar{\delta} = \frac{1}{K} \sum_{k=1}^K \delta_k$. This has expectation,

$$E \left[\hat{B}_\delta \right] = \pi_d^2 \left[\frac{\sigma^2 + n_d \bar{\mathbf{P}}_{a,d} \mathbf{V}_{a,o} \bar{\mathbf{P}}_{a,d}^T}{n_d} + \sigma_\delta^2 \right].$$

Which can also be expressed as,

$$E \left[\hat{B}_\delta \right] = E \left[\hat{B}_{DJ} \right] + \pi_d^2 \sigma_\delta^2.$$

The information anchored variance is,

$$E [V_{\text{anchored}}] = \frac{E [V_{\delta, \text{full}}] \left(E \left[\hat{W}_{DJ} \right] + E \left[\hat{B}_{DJ} \right] \right)}{E [V_{DJ, \text{full}}]} = E [V_{\delta, \text{full}}] \left[1 + \frac{E \left[\hat{B}_{DJ} \right]}{E \left[\hat{W}_{DJ} \right]} \right].$$

Since $E \left[\hat{W}_{DJ} \right] = E [V_{DJ, \text{full}}]$ this is,

$$E [V_{\text{anchored}}] = \mathbf{a}^T \mathbf{D}_{DJ} \Sigma \mathbf{D}_{DJ}^T \mathbf{a} + O \left(\frac{\pi_d(1-\pi_d)}{n} \right) + E \left[\hat{B}_{DJ} \right] + \frac{E \left[\hat{B}_{DJ} \right]}{E \left[\hat{W}_{DJ} \right]} O \left(\frac{\pi_d(1-\pi_d)}{n} \right).$$

If Rubin's variance estimator is information anchoring and preserves the information loss seen in the primary analysis under MAR then,

$$E \left[\hat{W}_\delta \right] + E \left[\hat{B}_\delta \right] \approx \mathbf{a}^T \mathbf{D}_{DJ} \Sigma \mathbf{D}_{DJ}^T \mathbf{a} + O \left(\frac{\pi_d(1-\pi_d)}{n} \right) + E \left[\hat{B}_{DJ} \right] + \frac{E \left[\hat{B}_{DJ} \right]}{E \left[\hat{W}_{DJ} \right]} O \left(\frac{\pi_d(1-\pi_d)}{n} \right).$$

That is,

$$\begin{aligned} \mathbf{a}^T \mathbf{D}_{DJ} \Sigma \mathbf{D}_{DJ}^T \mathbf{a} + O \left(\frac{\pi_d(1-\pi_d)}{n} \right) + E \left[\hat{B}_\delta \right] &\approx \mathbf{a}^T \mathbf{D}_{DJ} \Sigma \mathbf{D}_{DJ}^T \mathbf{a} + O \left(\frac{\pi_d(1-\pi_d)}{n} \right) + E \left[\hat{B}_{DJ} \right] \\ &+ \frac{E \left[\hat{B}_{DJ} \right]}{E \left[\hat{W}_{DJ} \right]} O \left(\frac{\pi_d(1-\pi_d)}{n} \right). \end{aligned}$$

After simplification and rearrangement this becomes,

$$0 \approx E \left[\hat{B}_{D,J} \right] - E \left[\hat{B}_\delta \right] + \frac{E \left[\hat{B}_{D,J} \right]}{E \left[\hat{W}_{D,J} \right]} O \left(\frac{\pi_d(1 - \pi_d)}{n} \right).$$

Which is,

$$0 \approx -\pi_d^2 \sigma_\delta^2 + \frac{E \left[\hat{B}_{D,J} \right]}{E \left[\hat{W}_{D,J} \right]} O \left(\frac{\pi_d(1 - \pi_d)}{n} \right). \quad (5.8)$$

These results confirm the observation made in Section 5.3, that the approximation between Rubin's variance estimator and the information anchored variance is principally driven by the variance of δ_k . If there is a great deal of uncertainty surrounding the proposed value of the difference between the deviators and those observed with the same history, i.e. large σ_δ^2 then the approximation will not be so sharp. We will lose information relative to the information anchored variance using Rubin's variance estimator. We note that when $\hat{B}_{D,J} \approx \hat{B}_\delta$ we will obtain a superior approximation of the ideal information anchored variance with Rubin's variance estimator. That is when σ_δ^2 is small.

Fixed delta

We now suppose $\delta_k \sim N(\delta, 0)$, i.e. δ is known in imputation. In this case we infer that $\hat{B}_{D,J} = \hat{B}_\delta$ since $\text{VAR}(x + \delta) = \text{VAR}(x)$ using standard variance sum rules. After simplification and rearrangement in this special case the criteria for information anchoring becomes,

$$0 \approx \frac{E \left[\hat{B}_{D,J} \right]}{E \left[\hat{W}_{D,J} \right]} O \left(\frac{\pi_d(1 - \pi_d)}{n} \right).$$

□

Since $\hat{B}_{D,J} = \hat{B}_\delta$, we obtain better agreement between Rubin's and the information anchored variance estimate when δ is fixed. The difference will be of a relatively small magnitude, in comparison to the information anchored variance. It consists of a quantity of $O(\pi_d(1 - \pi_d)n^{-1})$ multiplied by $E \left[\hat{B}_{D,J} \right] / E \left[\hat{W}_{D,J} \right]$. Since $E \left[\hat{B}_{D,J} \right]$ is typically much smaller than $E \left[\hat{W}_{D,J} \right]$ for realistic trial scenarios i.e. with small to moderate proportions of missing data as discussed in Section 4.1, the second term on the RHS will also be of a relatively small size to the information anchored variance, which is $O(n^{-1})$. To demonstrate this further we now conduct a simulation study.

5.5 Simulation study; baseline and single follow-up

This section reports the results of a simulation study conducted to assess the information anchoring properties of Rubin’s MI variance estimator following δ -adjustment, in a variety of trial scenarios with baseline (time 1) and a single follow-up (time 2). The baseline and 12 week FEV₁ data from the asthma trial formed the motivating example for generating the baseline and follow-up data. We consider simulations with a fixed and non-fixed δ -adjustment, where δ_k represents the change in response post-deviation for imputation k .

5.5.1 Methods

Baseline (time 1) and follow-up (time 2) data were generated using an underlying bivariate normal distribution with means and covariance matrices similar to those obtained in the asthma study at baseline and week 12 as,

$$\mu_{\text{placebo}} = [2, 1.9], \mu_{\text{active}} = [2, 2.2],$$

$$\Sigma_{\text{placebo}} = \Sigma_{\text{active}} = \begin{bmatrix} 0.4 & 0.2 \\ 0.2 & 0.6 \end{bmatrix}.$$

We consider scenarios with 10–70% missing data in the active arm only at time 2 where the unobserved data is postulated to be worse/greater by a fixed amount $\delta = \pm \{0, 0.1, 0.5, 1\}$. One thousand independent complete datasets were generated for each δ and missingness scenario, with a sample size of $n=500$ patients ($n=250$ placebo, $n=250$ active) for all simulations.

Interest lies in the mean treatment group difference at time 2. For each setting, the treatment effect (average treatment group difference in FEV₁ at time 2) and variance were computed using the full data. The analysis model was a linear regression of the time 2 response on randomised treatment group. Within the active arm, a MCAR missingness mechanism, was then imposed on the FEV₁ outcome at time 2. This was achieved by independently sampling a binary indicator from a Bernoulli distribution with success probability equal to the proportion of missing data required. If the indicator equalled one then the corresponding outcome at week 12 was set missing.

The ‘ δ -method’ outlined in Section 5.1 and in [44] was applied to each incomplete dataset with 50 imputations. Post-deviation data was regenerated using the appropriate conditional normal distributions to compute the treatment effect and its variance had we observed the post-deviation data under the particular δ setting ($V_{\delta, \text{full}}$).

The main outcomes of interest were Rubin’s MI estimate of variance for the treatment effect, and the ideal information anchored variance computed as detailed in Section 5.2. Estimates were averaged over the 1000 simulations in each scenario for comparison.

Additional simulations were run utilising the same set-up as above, but with a normally distributed

prior distribution on δ in imputation with mean values $\{0, -0.1, -0.5, -1\}$ and a variance of $\sigma_\delta^2 = 0.01^2$ or $\sigma_\delta^2 = 0.05^2$.

5.5.2 Results

In all scenarios the mean of Rubin's δ -adjusted treatment estimator over 1000 simulations was within 2 MCSE's of the true δ -adjusted treatment effect. Table 5.1 summarises the true treatment effect and the mean of Rubin's δ -adjusted treatment estimator ($\hat{\theta}_{\delta, \text{MI}}$) over 1000 simulations for $\delta_k = \{0, 0.1, 0.5, 1.0\}$. As expected, Rubin's treatment estimator is unbiased.

Missing Data	$\delta_k = 0$		$\delta_k = 0.1$		$\delta_k = 0.5$		$\delta_k = 1$	
	True θ	$\hat{\theta}_{\delta, \text{MI}}$	True θ	$\hat{\theta}_{\delta, \text{MI}}$	True θ	$\hat{\theta}_{\delta, \text{MI}}$	True θ	$\hat{\theta}_{\delta, \text{MI}}$
10%	0.300	0.297	0.310	0.307	0.350	0.348	0.400	0.401
20%	0.300	0.298	0.320	0.321	0.400	0.403	0.500	0.499
30%	0.300	0.297	0.330	0.330	0.450	0.447	0.600	0.601
40%	0.300	0.303	0.340	0.340	0.500	0.501	0.700	0.697
50%	0.300	0.301	0.350	0.350	0.550	0.549	0.800	0.802
60%	0.300	0.299	0.360	0.360	0.600	0.601	0.900	0.908
70%	0.300	0.302	0.370	0.365	0.650	0.647	1.000	1.003

Table 5.1: Rubin's MI treatment estimate ($\hat{\theta}_{\delta, \text{MI}}$, mean over 1000 simulations) vs. true treatment effect using the ' δ -method'.

We now consider the performance on Rubin's variance estimator. Over the range of fixed $\delta = \pm \{0, 0.1, 0.5, 1.0\}$ simulations show an excellent approximation of the information anchored variance by Rubin's variance estimator for all missingness scenarios, see Figures 5.1 and 5.2. For absolute values of δ_k up to and including 0.5 the information anchoring by Rubin's variance estimator looks practically perfect. It is only with a larger $\delta_k = \pm 1.0$ when very marginal differences start to appear for high proportions of missing data and Rubin's variance estimator slightly underestimates the information anchored variance. This is anticipated from the theory above. But the approximation is still excellent. We note that in the current setting $\delta_k = \pm 1.0$ is very large. The baseline data in both arms is generated with a mean of 2 and the response for the observed active cases at time 2 with mean 2.2. Thus $\delta_k = \pm 1.0$ corresponds to a postulated mean response of 3.2 or 1.2 at time 2 for the deviators.

Rubin's variance estimator is always greater than the variance estimate when the postulated δ scenario is true and the deviators are observed ($V_{\delta, \text{full}}$, labelled Deviators observed in Figures 5.1 and 5.2). Further Rubin's variance reflects the increased uncertainty for greater proportions of missing data.

In the δ -adjusted setting we also see that Rubin's variance estimator not only approximates the information anchored variance, but also corresponds with the empirical long-run sampling variance of the MI estimate (which is equivalent to the empirical long-run sampling variance of ML estimate) as expected. In each scenario, the computed sampling variance of the MI/ML estimates over the 1000 simulations, closely agrees with the average of Rubin's variance estimator. This is as expected since as discussed δ MI does not introduce an unwanted covariance between the patients of opposing treatment arms. This is unlike reference based MI where we borrow information from within the trial for imputation.

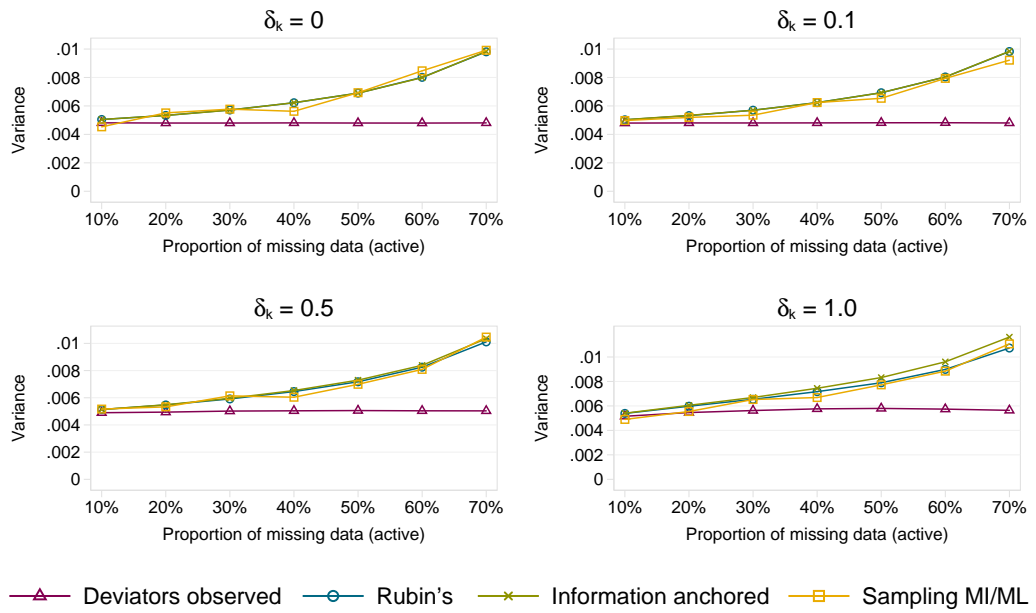


Figure 5.1: Rubin's $+\delta$ variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates. $\Delta = 0.3$, $n=250$ per arm.

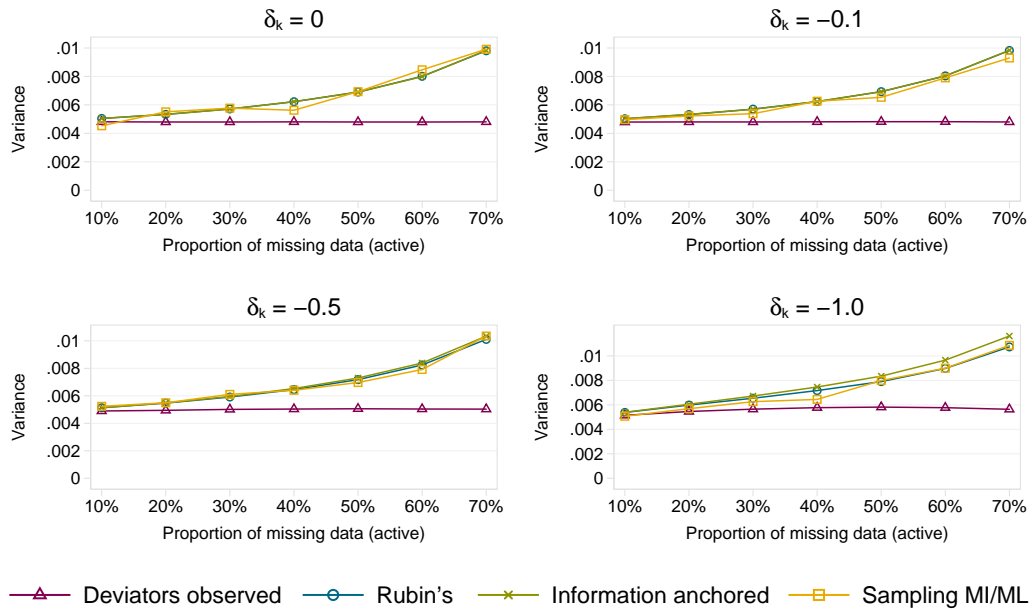


Figure 5.2: Rubin's $-\delta$ variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates. $\Delta = 0.3$, $n=250$ per arm.

Results from simulations with a normal prior distribution on δ_k in imputation are shown in Figures 5.3 to 5.4. As expected the approximation between Rubin's variance estimator and the information anchored variance is not quite as sharp when δ_k is no longer fixed. From Proposition 4, we expect an extra small component in the difference between Rubin's variance estimator and the information anchored variance which revolves around the proposed variance of δ_k . We indeed see the smaller the variance on δ_k the closer the approximation. In comparison to the fixed δ_k settings, with a larger

variance of 0.05^2 on δ_k . Rubin's slightly over estimates the required variance due to the additional variance on δ_k . This over estimation corresponds with Proposition 4. However Figure 5.4 indicates that overall we still see excellent approximation of the information anchored variance by Rubin's variance estimator with a variance of 0.05^2 . For larger variances of δ_k the approximation will not be so strong. When there is uncertainty in δ_k , we expect greater uncertainty surrounding our treatment estimate.

Although not shown here we also conducted simulations which examined the variance estimator for the baseline adjusted treatment effect, in the same settings as above. Results were the same, as anticipated following Proposition 4. Adjustment for baseline in the substantive model does not have a profound impact on the approximation.

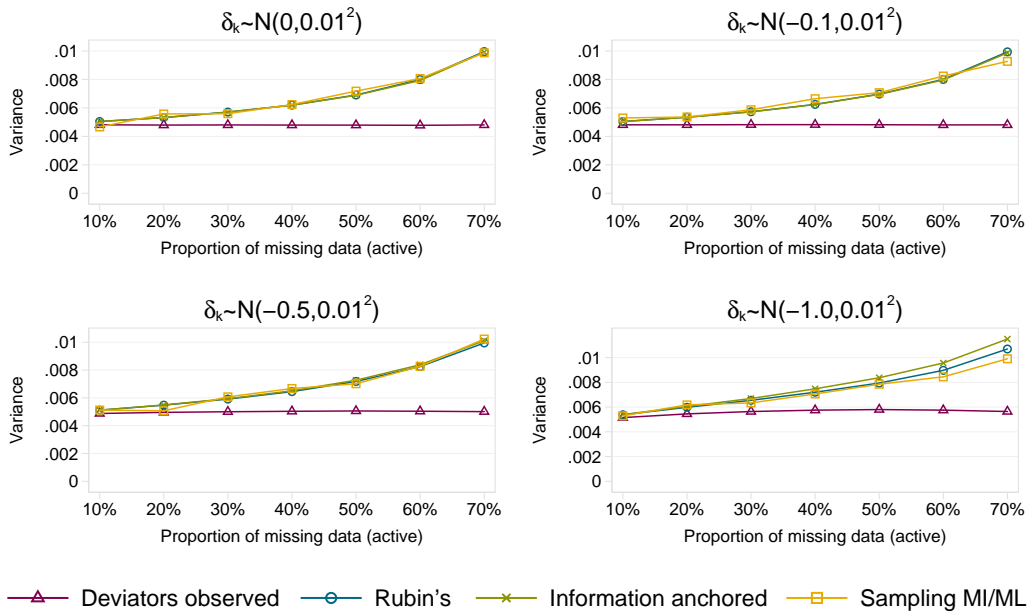


Figure 5.3: Rubin's $-\delta$ variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with $\sigma_\delta^2 = 0.01^2$. $\Delta = 0.3$, $n=250$ per arm.

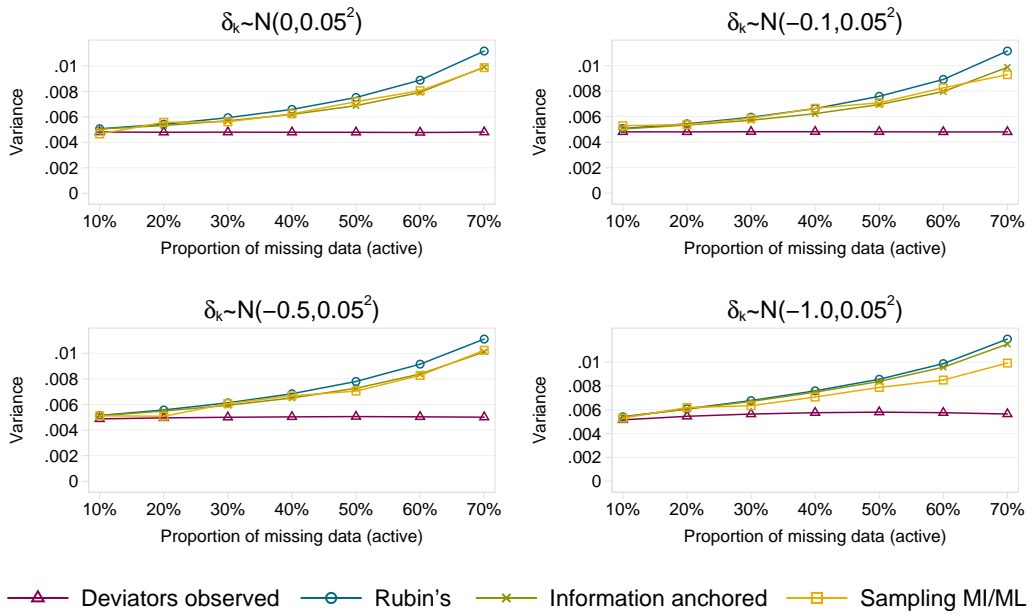


Figure 5.4: Rubin’s $-\delta$ variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with $\sigma_\delta^2 = 0.05^2$. $\Delta = 0.3$, $n=250$ per arm.

5.5.3 Discussion

The simulations confirm Rubin’s treatment estimator is unbiased and Proposition 4. That is, the negligible difference between Rubin’s variance estimator and the information anchored variance in the fixed δ -adjusted MI setting. Rubin’s variance estimator keeps the information constant across primary and sensitivity analysis. When δ is not fixed, the approximation will depend on the assumed variance for δ .

In the fixed δ -adjustment setting where $\sigma_\delta^2 = 0$, Rubin’s variance estimator and hence the information anchored variance match the empirical long-run sampling variance of the MI estimate (which is equivalent to that of the δ adjusted ML estimate). In this setting the sampling variance of the treatment estimate is not unnatural, since we do not impose a covariance between patients in opposing arms by borrowing within trial data for imputation. This provides further support for the use of MI with Rubin’s variance estimator in the reference based settings. Although as discussed the long-run empirical sampling variance is not correct in the reference based setting, since Rubin’s variance estimator has the same behaviour as in the δ -adjusted setting—in that it approximates the information anchored variance—we infer further justification.

We are not gaining or loosing any information relative to MAR, as desired, when using the fixed δ sensitivity analysis approach. When looking at differences in the mean response profiles for deviators and completers we want to keep the information constant.

5.6 General theory for Rubin’s variance estimator; longitudinal setting

We now extend Proposition 4 and present a general theory on Rubin’s variance estimator following δ -adjustment and how it anchors the information loss across the primary and sensitivity analysis for longitudinal trials. Our focus is on the treatment effect at the final time point. We also restrict our attention to monotone missingness patterns. As in the previous section, the proposition considers deviation in one arm only. The impact of deviation in both arm is explored in Section 5.8.

5.6.1 Setting for Proposition 5

Consider a two arm trial consisting of $2n$ patients. n patients have been randomised to an active arm and n patients to a reference arm, with a continuous outcome measured repeatedly over a series of visits, $j = 1, \dots, J$ (including baseline). The $(2n \times 1)$ vector of continuous outcome data at the final visit time J is denoted by \mathbf{Y} . A corresponding $(2n \times p)$ matrix \mathbf{x} , consists of the observed baseline covariates including baseline outcome and treatment. The estimated treatment effect can be written $\mathbf{a}^T \mathbf{Y}$ where \mathbf{a} may or may not incorporate \mathbf{x} .

We assume \mathbf{Y} is normally distributed and has known variance Σ . The variance of the data is assumed to be known without any serious loss of generality, following inference by Carpenter and Kenward in [44]. As discussed in Section 4.1 because of the conditional argument underlying the derivation this doesn’t have a profound impact on the exposition asymptotically.

Suppose all n reference patients are observed at baseline and up to time J without deviation. However among the n active treatment patients, only n_o patients are observed without deviation at all time points. Amongst the deviating n_d active arm patients, we observe $n_{d,j}$ patients who deviate in a monotone fashion following time j , for $j = 1, \dots, J - 1$. For simplicity we assume there is no interim missing data. Let \mathcal{D} and \mathcal{O} define the indices for the patients who do and do not deviate in the active arm respectively. Further $\mathcal{D}j$ denotes the index for deviating patients who deviate following time j , so that $n_d = \sum_{j=1}^{J-1} n_{d,j}$. $\pi_d = n_d/n$ denotes the total proportion of deviators in the active arm. $\pi_{d,j} = n_{d,j}/n$ denotes the proportion of active deviators who deviate following time j for $j = 1, \dots, J - 1$.

Typically, although not necessarily, some form of regression model will be the substantive model of interest, used to make inferences about the treatment effect in the population of interest. Appendix F.3 shows that,

$$V_{\delta, \text{full}} = V_{\text{DJ, full}} + Q, \tag{5.9}$$

where,

$$Q = \sum_{j=1}^{J-1} \frac{n_o n_{d,j} (J-j)^2 \delta^2}{n^3} + \sum_{p=1}^{J-1} \sum_{q=1}^{(J-1, q \neq p)} \frac{n_{d,p} n_{d,q} ((J-p)\delta - (J-q)\delta)^2}{n^3}, \quad (5.10)$$

for the mean treatment difference at time J , adjusted and unadjusted for baseline. We assume (5.9) holds and $\delta_k \sim N(\delta, \sigma_\delta^2)$.

5.6.2 Proposition 5

For sensitivity analyses of a two arm longitudinal trial with monotone missingness performed by MI using the ‘ δ -method’, where δ represents the change in rate of decline post-deviation as outlined in [44], the difference between Rubin’s variance estimator and the information anchoring variance for the treatment effect at time J is,

$$- \sum_{j=1}^{J-1} \pi_{d,j}^2 (J-j)^2 \sigma_\delta^2 + \frac{E[\hat{B}_{DJ}]}{E[\hat{W}_{DJ}]} Q.$$

That is,

$$\frac{V_{DJ, MI}}{V_{DJ, full}} = \frac{V_{\delta, MI}}{V_{\delta, full}} - \sum_{j=1}^{J-1} \pi_{d,j}^2 (J-j)^2 \sigma_\delta^2 + \frac{E[\hat{B}_{DJ}]}{E[\hat{W}_{DJ}]} Q.$$

When the change in rate of decline post-deviation is fixed, i.e. $\delta_k \sim N(\delta, 0)$ the difference between Rubin’s variance estimator and the information anchoring variance is,

$$\frac{E[\hat{B}_{DJ}]}{E[\hat{W}_{DJ}]} Q.$$

In other words,

$$\frac{V_{DJ, MI}}{V_{DJ, full}} = \frac{V_{\delta, MI}}{V_{\delta, full}} + \frac{E[\hat{B}_{DJ}]}{E[\hat{W}_{DJ}]} Q.$$

5.6.3 Proof of Proposition 5

Let $z = a, r$ index the randomised active arm or reference arm allocation for each patient i with follow-up outcome at time J denoted by Y_{ziJ} . The outcome data at the final time point for the reference patients are contained in the vector $\mathbf{Y}_{rJ} = (Y_{r1J}, \dots, Y_{rnJ})^T$. The final visit outcome data for the non-deviating active patients are contained in the vector $\mathbf{Y}_{aJ,o} = \{Y_{aiJ}; i \in \mathcal{O}\}^T$.

We suppose that each deviating patient has two potential outcomes at time J : the one that would occur if they remain on active treatment post-deviation (de-jure/MAR) and the other that would occur with δ -adjustment.

The potentially observable de-jure data for the n_d deviating patients at time J are contained in the vector $\mathbf{Y}_{aJ,DJ,d}$ and the alternative δ -adjusted outcome data in the vector $\mathbf{Y}_{aJ,\delta,d}$. Define $\mathbf{Y} = (\mathbf{Y}_{rJ}, \mathbf{Y}_{aJ,o}, \mathbf{Y}_{aJ,DJ,d}, \mathbf{Y}_{aJ,\delta,d})^T$ as the collection of observed and potentially observable outcome data, which has dimensions $[(n + n_o + 2n_d) \times 1]$.

For each deviating patient we can only observe one of the potential outcomes, either de-jure or δ -adjusted. Consider two $[(n + n_o + 2n_d) \times (n + n_o + 2n_d)]$ matrices, \mathbf{D}_{DJ} and \mathbf{D}_δ of 0's and 1's, arranged such that $\mathbf{D}_{DJ}\mathbf{Y}$ gives the $[(n + n_o + 2n_d) \times 1]$ de-jure data and $\mathbf{D}_\delta\mathbf{Y}$ gives the $[(n + n_o + 2n_d) \times 1]$ δ -adjusted data at time J .

Let \mathbf{a} be a $[(n + n_o + 2n_d) \times 1]$ vector such that $\mathbf{a}^T\mathbf{D}_{DJ}\mathbf{Y}$ returns the de-jure treatment estimate and $\mathbf{a}^T\mathbf{D}_\delta\mathbf{Y}$ returns the treatment estimate, with unobserved outcomes adjusted by δ . When the deviating patients experience de-jure behaviour post-deviation and are observed the expectation of the variance of the de-jure on-treatment estimand can be expressed as in (5.6). We consider settings where the expectation of the variance of our δ -adjusted estimand can then be expressed as in (5.7).

We now suppose that post-deviation data are unobserved, i.e. the potentially observable de-jure and δ -adjusted entries in \mathbf{Y} are missing for the n_d active patients. We alternatively impute these outcomes, using de-jure imputation and δ -adjusted imputation. This gives K 'complete' data samples \mathbf{Y}_k , of size $[(n + n_o + 2n_d) \times 1]$. For this we need appropriate imputation distributions for each missing data pattern under each scenario, with suitable posteriors for the included parameters.

Under our de-jure assumption (on-treatment MAR), the imputation model for patients deviating following time j , for $j = 1, \dots, J - 1$ is formed from the regression of $\mathbf{Y}_{aJ,o}$ on $\mathbf{P}_{a,o,j}$ where $\mathbf{P}_{a,o,j}$ is the design matrix for the imputation model, which contains the values of the $1, \dots, j$ outcomes and covariates included in the imputation model for the n_o observed active patients, along with a vector of 1's to include an intercept in the model. This forms the required imputation models since we are only considering monotone missing patterns and interested in the treatment effect at time J . We do not need to impute any missing outcomes before time J .

The parameter estimates for the de-jure imputation model for the $n_{d,j}$ patients missing outcomes $j + 1$, to J are found using $\hat{\beta}_{a,o,j} = (\mathbf{P}_{a,o,j}^T \mathbf{P}_{a,o,j})^{-1} \mathbf{P}_{a,o,j}^T \mathbf{Y}_{aJ,o}$ with known covariance matrix $\mathbf{V}_{a,o,j} = (\mathbf{P}_{a,o,j}^T \mathbf{P}_{a,o,j})^{-1} \sigma_j^2$. We assume the large sample posterior for the parameter estimates of the de-jure imputation model, denoted $\hat{\beta}_{DJ,j}$, is normal and centered on the ML estimator $\hat{\beta}_{a,o,j}$ with known covariance matrix $\mathbf{V}_{a,o,j}$. That is,

$$\hat{\boldsymbol{\beta}}_{DJ,j} | Y_{aJ,o} \sim N(\hat{\boldsymbol{\beta}}_{a,o,j}; \mathbf{V}_{a,o,j}).$$

The de-jure imputation model for active patient i deviating following time j , for $j = 1, \dots, J - 1$ and imputation k can therefore be expressed as,

$$\tilde{Y}_{aiJ,k} | \mathbf{Y}_{aJ,o} = \mathbf{P}_{a,d,j,i} \left[\hat{\boldsymbol{\beta}}_{a,o,j} + \mathbf{b}_{a,o,j,k} \right] + e_{i,j,k} \text{ for } i \in \mathcal{DJ},$$

where, $\mathbf{b}_{a,o,j,k} \sim N(0, \mathbf{V}_{a,o,j})$, $e_{i,j,k} \sim N(0, \sigma_j^2)$ and $\mathbf{P}_{a,d,j,i}$ contains the values of the $1, \dots, j$ outcomes and covariates included in the imputation model (excluding treatment plus a 1 for the intercept) for each deviating patient i , who deviates following time j .

For δ -adjusted imputation we must add $(J - j)\delta_k$ into the imputation model. The δ imputation model for active patients deviating following time j , for $j = 1, \dots, J$ can be expressed as,

$$\tilde{Y}_{aiJ,k} | \mathbf{Y}_{aJ,o} = \mathbf{P}_{a,d,j,i} \left[\hat{\boldsymbol{\beta}}_{a,o,j} + \mathbf{b}_{a,o,j,k} \right] + (J - j)\delta_k + e_{i,j,k} \text{ for } i \in \mathcal{DJ},$$

where, $\mathbf{b}_{a,o,j,k} \sim N(0, \mathbf{V}_{a,o,j})$, $\delta_k \sim N(\delta, \sigma_\delta^2)$ and $e_{i,j,k} \sim N(0, \sigma_j^2)$. We are interested in imputation inference for, $\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_{DJ} \mathbf{Y}_k$ or $\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_\delta \mathbf{Y}_k$. For Rubin's variance estimator, under the conditions of the theorem,

$$E \left[\hat{W}_{DJ} \right] = E \left[\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_{DJ} \hat{\boldsymbol{\Sigma}}_k \mathbf{D}_{DJ}^T \mathbf{a} \right] \rightarrow \mathbf{a}^T \mathbf{D}_{DJ} \boldsymbol{\Sigma} \mathbf{D}_{DJ}^T \mathbf{a},$$

and,

$$E \left[\hat{W}_\delta \right] = E \left[\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_\delta \hat{\boldsymbol{\Sigma}}_k \mathbf{D}_\delta^T \mathbf{a} \right] \rightarrow \mathbf{a}^T \mathbf{D}_{DJ} \boldsymbol{\Sigma} \mathbf{D}_{DJ}^T \mathbf{a} + Q.$$

Under de-jure,

$$\hat{B}_{DJ} = \frac{1}{K-1} \sum_{k=1}^K \left[\sum_{j=1}^{J-1} (\pi_{d,j} (\bar{e}_{j,k} - \bar{e}_j) + \pi_{d,j} (\bar{\mathbf{P}}_{\mathbf{a},d,j} \mathbf{b}_{\mathbf{a},o,j,k} - \bar{\mathbf{P}}_{\mathbf{a},d,j} \bar{\mathbf{b}}_{\mathbf{a},o,j})) \right]^2.$$

where $\bar{e}_{j,k} = \frac{1}{n_{d,j}} \sum_{i \in \mathcal{D}j} e_{i,j,k}$, $\bar{e}_j = \frac{1}{K} \sum_{k=1}^K \bar{e}_{j,k}$, $\bar{\mathbf{P}}_{a,d,j} = \frac{1}{n_{d,j}} \sum_{i \in \mathcal{D}j} \mathbf{P}_{a,d,j,i}$ and $\bar{b}_{a,o,j} = \frac{1}{K} \sum_{k=1}^K b_{a,o,j,k}$. Which has expectation,

$$E \left[\hat{B}_{\text{DJ}} \right] = \sum_{j=1}^{J-1} \pi_{d,j}^2 \left[\frac{\sigma_j^2 + n_{d,j} \bar{\mathbf{P}}_{a,d,j} \mathbf{V}_{a,o,j} \bar{\mathbf{P}}_{a,d,j}^T}{n_{d,j}} \right].$$

Under δ -adjustment,

$$\hat{B}_\delta = \frac{1}{K-1} \sum_{k=1}^K \left[\sum_{j=1}^{J-1} (\pi_{d,j} (\bar{e}_{j,k} - \bar{e}_j) + \pi_{d,j} (\bar{\mathbf{P}}_{a,d,j} \mathbf{b}_{a,o,j,k} - \bar{\mathbf{P}}_{a,d,j} \bar{\mathbf{b}}_{a,o,j}) + \pi_{d,j} (J-j) (\delta_k - \bar{\delta})) \right]^2$$

where $\bar{\delta} = \frac{1}{K} \sum_{k=1}^K \delta_k$. Which has expectation,

$$E \left[\hat{B}_\delta \right] = \sum_{j=1}^{J-1} \pi_{d,j}^2 \left[\frac{\sigma_j^2 + n_{d,j} \bar{\mathbf{P}}_{a,d,j} \mathbf{V}_{a,o,j} \bar{\mathbf{P}}_{a,d,j}^T + n_{d,j} (J-j)^2 \sigma_\delta^2}{n_{d,j}} \right].$$

This can also be expressed as,

$$E \left[\hat{B}_\delta \right] = E \left[\hat{B}_{\text{DJ}} \right] + \sum_{j=1}^{J-1} \pi_{d,j}^2 (J-j)^2 \sigma_\delta^2.$$

The information anchored variance is,

$$E [V_{\text{anchored}}] = \mathbf{a}^T \mathbf{D}_{\text{DJ}} \boldsymbol{\Sigma} \mathbf{D}_{\text{DJ}}^T \mathbf{a} + O \left(\frac{\pi_d (1 - \pi_d)}{n} \right) + E \left[\hat{B}_{\text{DJ}} \right] + \frac{E \left[\hat{B}_{\text{DJ}} \right]}{E \left[\hat{W}_{\text{DJ}} \right]} Q.$$

If Rubin's variance estimator is information anchoring then,

$$\mathbf{a}^T \mathbf{D}_{\text{DJ}} \boldsymbol{\Sigma} \mathbf{D}_{\text{DJ}}^T \mathbf{a} + Q + E \left[\hat{B}_\delta \right] \approx \mathbf{a}^T \mathbf{D}_{\text{DJ}} \boldsymbol{\Sigma} \mathbf{D}_{\text{DJ}}^T \mathbf{a} + O \left(\frac{\pi_d (1 - \pi_d)}{n} \right) + E \left[\hat{B}_{\text{DJ}} \right] + \frac{E \left[\hat{B}_{\text{DJ}} \right]}{E \left[\hat{W}_{\text{DJ}} \right]} Q.$$

After simplification and rearrangement this becomes,

$$0 \approx E \left[\hat{B}_{Dj} \right] - E \left[\hat{B}_\delta \right] + \frac{E \left[\hat{B}_{Dj} \right]}{E \left[\hat{W}_{Dj} \right]} Q,$$

which is,

$$0 \approx - \sum_{j=1}^{J-1} \pi_{d,j}^2 (J-j)^2 \sigma_\delta^2 + \frac{E \left[\hat{B}_{Dj} \right]}{E \left[\hat{W}_{Dj} \right]} Q.$$

Similar to the baseline and single follow-up setting we see the approximation between Rubin's variance estimator and the information anchored variance is principally driven by the variance of δ_k . If there is a great deal of uncertainty surrounding the difference in the decline/improvement in response between the deviators and those observed with the same history, i.e. large σ_δ^2 then the approximation will not be so sharp. Specifically with longitudinal data, we see the impact of the σ_δ^2 also depends on observed deviation patterns, since the σ_δ^2 term is multiplied by $(J-j)^2$. If a number of patients deviate early in the trial there will be a greater difference between Rubin's variance estimator and the information anchored variance, in comparison to similar settings with deviations observed nearer the end of follow-up. If patients deviate later in the trial the impact of σ_δ^2 will not be as great. If we are more uncertain about the rate of change in response post-deviation, it is logical that there will be more uncertainty in our treatment estimate, thus the approximation wont be as strong.

We additionally note here that if one applies a blanket δ -adjustment to all imputed outcomes, regardless of the time since deviation (rather than incorporating δ as a rate of change) then the result becomes,

$$0 \approx - \sum_{j=1}^{J-1} \pi_{d,j}^2 \sigma_\delta^2 + \frac{E \left[\hat{B}_{Dj} \right]}{E \left[\hat{W}_{Dj} \right]} Q.$$

The approximation does not depend on the initial time of deviation in the same way. This result is a natural extension of the result in the baseline and single-follow up setting described in Section 5.4.

For both blanket δ adjustment and where δ represents a rate of change in response, the second term in the difference consists of a quantity which is Q (5.10) multiplied by $E \left[\hat{B}_{Dj} \right] / E \left[\hat{W}_{Dj} \right]$. Q is a term that is $O(\pi_{d,j}(1 - \pi_d)n^{-1}) + O(\pi_{d,p}\pi_{d,p}n^{-1})$. The second term therefore has a small magnitude since, as previously discussed in Section 4.1, $E \left[\hat{B}_{Dj} \right]$ is typically smaller than $E \left[\hat{W}_{Dj} \right]$. The information anchored variance we are seeking is much larger in comparison. This is $O(n^{-1})$.

Fixed delta

Suppose $\delta_k \sim N(\delta, 0)$. In this case we infer that $\hat{B}_{D,J} = \hat{B}_\delta$ since $\text{VAR}(x + \delta) = \text{VAR}(x)$ (i.e. using standard variance sum rules). Since $\hat{B}_{D,J} = \hat{B}_\delta$ we obtain a better approximation of the information anchored variance by Rubin's variance estimator when δ is fixed. Whether we adjust by a simple δ at all time points or multiples of δ to represent a change in rate of decline post-deviation, when δ is fixed we get,

$$0 \approx \frac{E[\hat{B}_{D,J}]}{E[\hat{W}_{D,J}]}Q.$$

□

There is improved information anchoring by Rubin's variance when δ is fixed. Following the above inferences the difference will be of a practically negligible size. We now conduct a simulation study to confirm this.

5.7 Simulation study; longitudinal setting

Here we present the results of a simulation study, performed to evaluate the information anchoring properties of Rubin's MI variance estimator in the δ -adjustment sensitivity analysis setting, with a continuous baseline (time 1) and two follow-up outcomes (time 2 and time 3) subject to non-response. The baseline, week 4 and week 12 data from the asthma trial (described in Subsection 1.7.2) formed the motivating example for generating the datasets. We consider simulations with a fixed and non-fixed δ -adjustment, where δ_k represents the change in rate of response post-deviation for imputation k . Therefore for patients who deviated following baseline their MAR imputed observation at time 2 is altered by δ_k and at time 3 by $2\delta_k$ for imputation k . For patients who deviated following time 2 their MAR imputed observation at time 3 is altered by δ_k .

5.7.1 Methods

Baseline (time 1), time 2 and time 3 data were generated using an underlying MVN distribution with means and covariance matrices similar to those obtained in the asthma trial at baseline, week 4 and week 12 as,

$$\mu_{\text{placebo}} = [2, 1.95, 1.9], \mu_{\text{active}} = [2, 2.21, 2.2],$$

$$\Sigma_{\text{placebo}} = \Sigma_{\text{active}} = \begin{bmatrix} 0.4 & 0.2 & 0.2 \\ 0.2 & 0.5 & 0.2 \\ 0.2 & 0.2 & 0.6 \end{bmatrix}.$$

We consider a range of scenarios with 5–35% of active patients missing data at week 4 and week 12 and a further 5–35% of active patients missing week 12 only, corresponding to a total of 10–70% missing data at week 12 in the active arm. Table E.1 summarises the specific proportions of missing data at each time point in each scenario. The unobserved data is postulated to be worse/greater by a fixed postulated amount δ which takes values $\pm \{0, 0.1, 0.5, 1\}$. One thousand independent complete datasets were generated for each δ and missingness scenario, and a sample size of $n=500$ patients ($n=250$ placebo, $n=250$ active) was drawn for all simulations.

The analysis model was a linear regression of the time 3 response on randomised treatment group. The key parameter of interest was the the mean treatment group difference at time 3. For each setting, the treatment effect (average treatment group difference in FEV₁ at time 3) and variance were computed using the full data.

A MCAR monotone missingness pattern was then imposed within the active arm by independently sampling a binary indicator from a Bernoulli distribution with success probability equal to the proportion of missing data required at time 2 for each active patient. If the indicator equalled one at time 2 then the corresponding outcomes at time 2 and time 3 were set missing. A second binary indicator was independently sampled for active patients observed at time 2 from a Bernoulli distribution with success probability equal to the proportion of additional missing data required at time 3 to obtain the second missingness pattern.

The ‘ δ -method’ outlined in Section 5.1 and in [44] was applied to each incomplete dataset with 50 imputations. Post-deviation data were regenerated using the appropriate conditional normal distributions so that the treatment effect and its variance when we observe post-deviation data under the particular δ scenario could be computed ($V_{\delta, \text{full}}$).

The main outcomes of interest were Rubin’s MI estimate of variance for the treatment effect, and the ideal information anchored variance computed as in Section 5.2. Estimates were averaged over the 1000 simulations in each scenario for comparison. Subsequent simulations were run utilising the same set-up as above, but with a normal prior distribution on δ in imputation, with mean values $\{0, -0.1, -0.5, -1\}$ and a variance of $\sigma_{\delta}^2 = 0.01^2$ or $\sigma_{\delta}^2 = 0.05^2$.

5.7.2 Results

Simulations show excellent information anchoring by Rubin’s variance estimator for all missingness scenarios, over the range of fixed delta ($\delta_k = \pm \{0, 0.1, 0.5\}$), see Figures 5.5 and 5.6. For large $\delta_k = \pm 1.0$ the approximation is excellent for up to 40–50% missingness. For greater amounts of missingness the approximation is not quite as strong, but still good. In this setting, for patients who deviated following time 2, their MAR imputed outcome at the second follow-up (time 3) is adjusted by $\delta_k = \pm 1.0$. For patients deviating post-baseline, their MAR imputed outcome at the second follow-up time point (time 3) is adjusted by $2\delta_k = \pm 2.0$. We therefore note that this adjustment is very large and highly unrealistic. $\delta_k = \pm 1.0$ is an extreme case scenario corresponding to a postulated mean response of 4.2 or 0.2 at week 12 for patients who deviated following baseline. In practice we would not anticipate such big δ_k . Here a large value still shows good agreement indicating the appropriateness of Rubin’s variance estimator.

In the longitudinal δ -adjusted setting we also see that Rubin’s variance estimator corresponds with

the empirical long-run sampling variance of the MI estimate (which corresponds the empirical long-run sampling variance of the ML estimate) as expected.

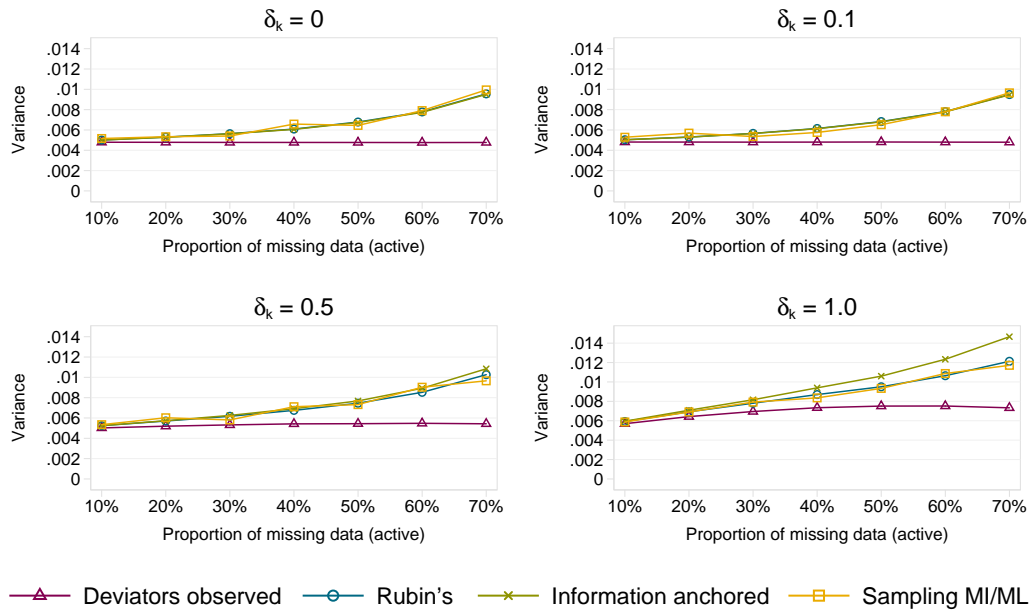


Figure 5.5: Rubin's $+\delta$ variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, in the longitudinal setting. $\Delta = 0.3$, $n=250$ per arm.

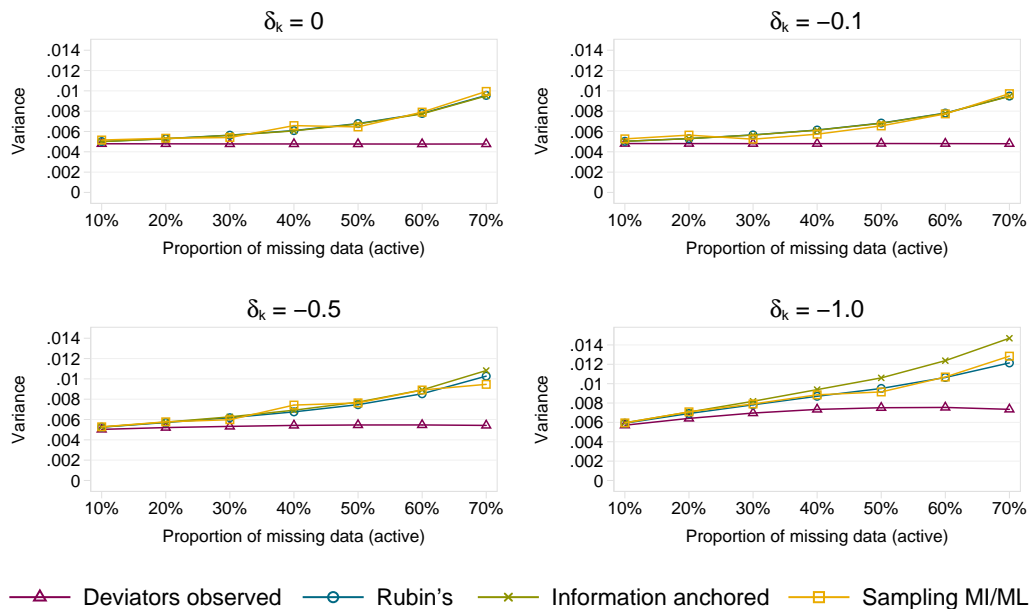


Figure 5.6: Rubin's $-\delta$ variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, in the longitudinal setting. $\Delta = 0.3$, $n=250$ per arm.

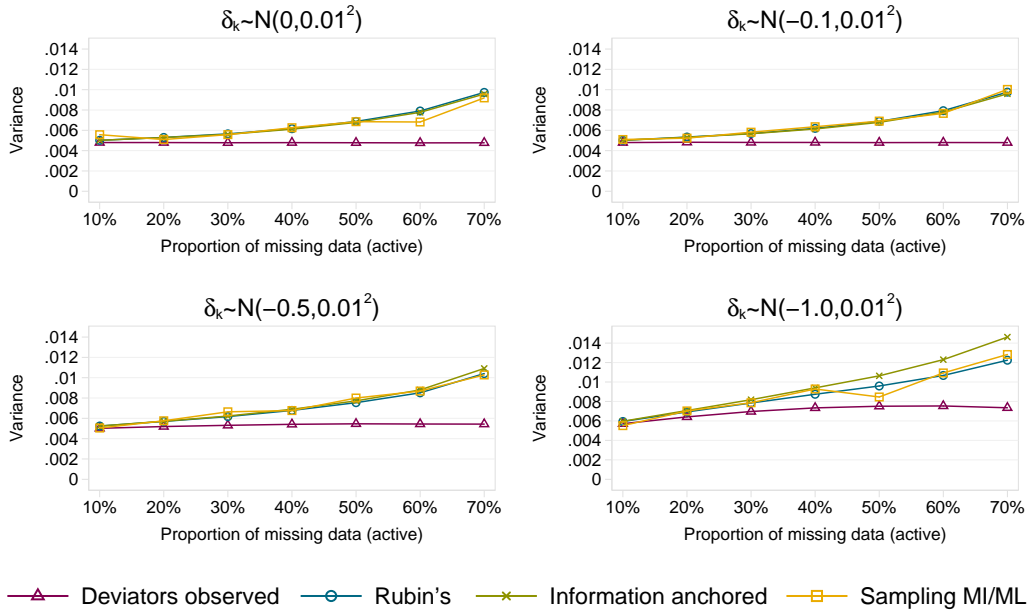


Figure 5.7: Rubin's $-\delta$ variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, in the longitudinal setting, with $\sigma_\delta^2 = 0.01^2$. $\Delta = 0.3$, $n=250$ per arm.

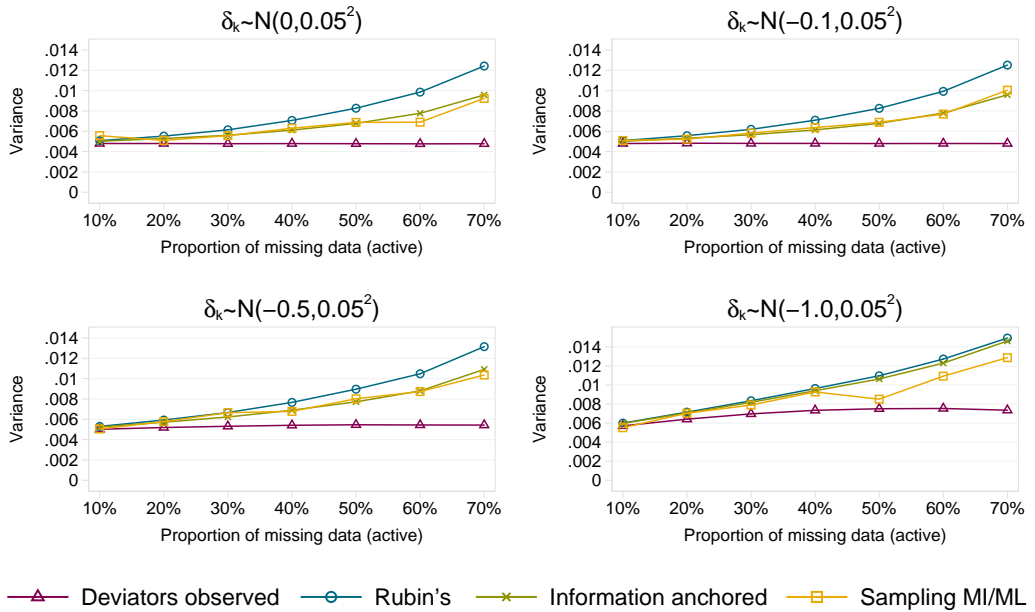


Figure 5.8: Rubin's $-\delta$ variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, in the longitudinal setting, with $\sigma_\delta^2 = 0.05^2$. $\Delta = 0.3$, $n=250$ per arm.

As expected, the agreement between Rubin's variance estimator and the information anchored variance is better when δ_k is fixed. Results from simulations with a prior distribution on δ_k are shown in Figures 5.7 and 5.8. We see that generally, the smaller the variance of δ_k the stronger the approximation, corresponding with Proposition 5. Unless a very large unrealistic $\delta_k = 1.0$ is used. In this case we see better agreement between Rubin's variance formula and the information anchored variance with $\sigma_\delta^2 = 0.05^2$ than with $\sigma_\delta^2 = 0.01^2$. In this setting the impact of the large

δ_k happens to balance with the impact of $\sigma_\delta^2 = 0.05^2$. Proposition 5 indicates the variability of δ_k , σ_δ^2 , has a negative impact on the difference, whilst the magnitude of δ_k has a positive impact. Generally we lose information relative to the information anchored variance using Rubin’s variance estimator when σ_δ^2 is large.

Although not shown here we also conducted simulations where the baseline adjusted treatment effect was of interest, for which results were the same. In all simulation settings, as expected, Rubin’s treatment estimator was additionally confirmed as unbiased (within 2 MCSE’s of the true δ adjusted treatment effect).

5.7.3 Discussion

The simulation results show that Rubin’s variance estimator keeps the information constant across primary and sensitivity analysis in realistic longitudinal trial settings with monotone missingness patterns and fixed δ adjustment. The results correspond with the theory outlined in Section 5.6. The information anchoring performance of Rubin’s variance estimator is excellent, even with high proportions of missing data.

The desired information anchored variance, which matches the empirical long-run sampling variance more closely, is achieved using Rubin’s variance estimator. This supports the use of Rubin’s variance estimator not only in the δ -adjusted MI setting, but additionally in the reference based MI setting, where the information anchored variance is also obtained using Rubin’s variance estimator.

If δ is not fixed and specifically there is a large variance on δ_k , i.e. there is a great deal of uncertainty surrounding the difference in outcome between those who deviated and those who do not with similar history then the approximation of the information anchored variance will not be as strong. Rubin’s variance estimator will lose information relative to the information anchored variance. Trialists should therefore carefully consider the proposed variance for δ a-priori.

5.8 Extension for deviation in both arms

The proofs of Proposition 4 and 5 show Rubin’s variance estimator approximates the information anchored variance in the δ -adjustment sensitivity analysis setting with deviation in one arm only. We now explore the impact of deviation in both arms and show how the results hold with deviation in both arms. With deviation in both arms we can apply the same δ -adjustment for all deviating patients, regardless of their randomised treatment arm. Alternatively we can vary the δ -adjustment by trial arm, that is adjust using $\delta_z \sim N(\delta_{z,k}, \sigma_{\delta_z}^2)$ for treatment arm z and imputation k .

We first re-consider the baseline and single follow-up setting in Proposition 4, but now with deviation in both arms where the same δ -adjustment is made, regardless of the patients randomised arm. We assume $\delta_z \sim N(\delta_k, \sigma_\delta^2)$ for $z = a, r$.

Suppose in addition to the conditions already laid down in Proposition 4, among the n reference patients, only $n_{r,o}$ are actually observed, whilst the remaining $n_{r,d}$ deviate post-baseline. Let \mathcal{RD} and \mathcal{RO} define the sets of indices for patients who do and do not deviate in the reference arm re-

spectively. $\pi_{r,d}$ denotes the proportion of deviators in the reference arm, i.e. $n_{r,d}/n$. The outcome data for the observed reference patients are contained in the vector $\mathbf{Y}_{r,o} = \{Y_{ri}; i \in \mathcal{RO}\}^T$. The potentially observable de-jure data for the $n_{r,d}$ deviating reference patients are contained in the vector $\mathbf{Y}_{r,DJ,d}$ and the alternative δ -adjusted outcome data in the vector $\mathbf{Y}_{r,\delta,d}$. The full collection of observed and potentially observable outcome data is now defined as,

$\mathbf{Y} = (\mathbf{Y}_{r,o}, \mathbf{Y}_{r,DJ,d}, \mathbf{Y}_{r,\delta,d}, \mathbf{Y}_{a,o}, \mathbf{Y}_{a,DJ,d}, \mathbf{Y}_{a,\delta,d})^T$ which has dimensions $[(n_{r,o} + 2n_{r,d} + n_o + 2n_d)]$. We assume \mathbf{Y} is normally distributed and has known variance Σ .

We redefine the two matrices \mathbf{D}_{DJ} and \mathbf{D}_δ so that $\mathbf{D}_{DJ}\mathbf{Y}$ and $\mathbf{D}_\delta\mathbf{Y}$ now each give the de-jure data or δ -adjusted data across both treatment arms.

When the treatment effect of interest is the mean difference in the follow-up outcome, Appendix B.1 shows that with deviation in both arms,

$$E[V_{\delta, \text{full}}] = \mathbf{a}^T \mathbf{D}_{DJ} \Sigma \mathbf{D}_{DJ}^T \mathbf{a} + O\left(\frac{\pi_d(1-\pi_d)}{n}\right) + O\left(\frac{\pi_{r,d}(1-\pi_{r,d})}{n}\right). \quad (5.11)$$

We focus this exposition on settings where (5.11) holds. Following the inference in Section 3.4, this will also be true for the baseline adjusted treatment estimator. We follow the same developments as in Section 5.4 to establish our de-jure (on-treatment MAR) and δ -adjusted imputation models for the deviating active arm patients. For the reference arm, under our de-jure assumption the imputation model is formed from the regression of $\mathbf{Y}_{r,o}$ on $\mathbf{P}_{r,o}$ where $\mathbf{P}_{r,o}$ is the $[n_{r,o} \times p]$ design matrix for the imputation model, which contains the values of the $(p-1)$ covariates included in the imputation model (including the baseline outcome but excluding the treatment indicator since we perform imputation separately by arm) with a vector of 1's to include an intercept term in the model, for observed reference patients.

The parameter estimates for the de-jure imputation model, denoted $\hat{\beta}_{DJ,r}$, are found using, $\hat{\beta}_{r,o} = (\mathbf{P}_{r,o}^T \mathbf{P}_{r,o})^{-1} \mathbf{P}_{r,o}^T \mathbf{Y}_{r,o}$ with known covariance matrix $\mathbf{V}_{r,o} = (\mathbf{P}_{r,o}^T \mathbf{P}_{r,o})^{-1} \sigma^2$. We assume the large sample posterior for $\hat{\beta}_{DJ,r}$ is normal and centered on the ML estimator $\hat{\beta}_{r,o}$ with covariance matrix $\mathbf{V}_{r,o}$, that is,

$$\hat{\beta}_{DJ} | \mathbf{Y}_{r,o} \sim N(\hat{\beta}_{r,o}; \mathbf{V}_{r,o}).$$

The de-jure imputation model for patient i and imputation k in the reference arm can therefore be expressed as,

$$\tilde{Y}_{ri,k} | \mathbf{Y}_{r,o} = \mathbf{P}_{r,d,i} \left[\hat{\beta}_{r,o} + \mathbf{b}_{r,o,k} \right] + e_{r,i,k} \text{ for } i \in \mathcal{RD},$$

where $\mathbf{b}_{r,o,k} \sim N(0, \mathbf{V}_{r,o})$, $e_{r,i,k} \sim N(0, \sigma^2)$ and $\mathbf{P}_{r,d,i}$ is the covariate data for each deviating reference patient i (excluding treatment group but including a 1), of dimensions $[1 \times p]$.

For δ -adjusted imputation we must add δ_k into the imputation model. The δ imputation model for patient i and imputation k in the reference arm can therefore be expressed as,

$$\check{Y}_{r,i,k} | \mathbf{Y}_{r,o} = \mathbf{P}_{r,d,i} \left[\hat{\boldsymbol{\beta}}_{r,o} + \mathbf{b}_{r,o,k} \right] + \delta_k + e_{r,i,k} \text{ for } i \in \mathcal{RD},$$

where additionally $\delta_k \sim N(\delta, \sigma_\delta^2)$. We are interested in imputation inference for, $\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_{\text{DJ}} \mathbf{Y}_k$ or $\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_\delta \mathbf{Y}_k$. Letting the number of imputations, $K \rightarrow \infty$, the variance of our MI treatment estimate as estimated by Rubin's rules is, $V_{\text{DJ, MI}} = \hat{W}_{\text{DJ}} + \hat{B}_{\text{DJ}}$ or $V_{\delta, \text{MI}} = \hat{W}_\delta + \hat{B}_\delta$ where under the conditions described,

$$E \left[\hat{W}_{\text{DJ}} \right] = E \left[\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_{\text{DJ}} \hat{\boldsymbol{\Sigma}}_k \mathbf{D}_{\text{DJ}}^T \mathbf{a} \right] \rightarrow \mathbf{a}^T \mathbf{D}_{\text{DJ}} \boldsymbol{\Sigma} \mathbf{D}_{\text{DJ}}^T \mathbf{a},$$

and,

$$E \left[\hat{W}_\delta \right] = E \left[\frac{1}{K} \sum_{k=1}^K \mathbf{a}^T \mathbf{D}_\delta \hat{\boldsymbol{\Sigma}}_k \mathbf{D}_\delta^T \mathbf{a} \right] \rightarrow \mathbf{a}^T \mathbf{D}_{\text{DJ}} \boldsymbol{\Sigma} \mathbf{D}_{\text{DJ}}^T \mathbf{a} + O \left(\frac{\pi_d(1-\pi_d)}{n} \right) + O \left(\frac{\pi_{r,d}(1-\pi_{r,d})}{n} \right).$$

Under de-jure,

$$\begin{aligned} \hat{B}_{\text{DJ}} = \frac{1}{K-1} \sum_{k=1}^K & \left[\pi_d (\bar{e}_k - \bar{e}) + \pi_d (\bar{\mathbf{P}}_{a,d} \mathbf{b}_{a,o,k} - \bar{\mathbf{P}}_{a,d} \bar{\mathbf{b}}_{a,o}) - \pi_{r,d} (\bar{e}_{r,k} - \bar{e}_r) \right. \\ & \left. - \pi_{r,d} (\bar{\mathbf{P}}_{r,d} \mathbf{b}_{r,o,k} - \bar{\mathbf{P}}_{r,d} \bar{\mathbf{b}}_{r,o}) \right]^2, \end{aligned}$$

where $\bar{e}_{r,k} = \frac{1}{n_{r,d}} \sum_{i \in \mathcal{RD}} e_{r,i,k}$, $\bar{e}_r = \frac{1}{K} \sum_{k=1}^K \bar{e}_{r,k}$, $\bar{\mathbf{P}}_{r,d} = \frac{1}{n_{r,d}} \sum_{i \in \mathcal{RD}} \mathbf{P}_{r,d,i}$ and $\bar{\mathbf{b}}_{r,o} = \frac{1}{K} \sum_{k=1}^K \mathbf{b}_{r,o,k}$. This has expectation,

$$E \left[\hat{B}_{\text{DJ}} \right] = \pi_d^2 \left[\frac{\sigma^2 + n_d \bar{\mathbf{P}}_{a,d} \mathbf{V}_{a,o} \bar{\mathbf{P}}_{a,d}^T}{n_d} \right] + \pi_{r,d}^2 \left[\frac{\sigma^2 + n_{r,d} \bar{\mathbf{P}}_{r,d} \mathbf{V}_{r,o} \bar{\mathbf{P}}_{r,d}^T}{n_{r,d}} \right].$$

With δ -adjustment,

$$\hat{B}_\delta = \frac{1}{K-1} \sum_{k=1}^K \left[\pi_d (\bar{e}_k - \bar{e}) + \pi_d (\bar{\mathbf{P}}_{a,d} \mathbf{b}_{a,o,k} - \bar{\mathbf{P}}_{a,d} \bar{\mathbf{b}}_{a,o}) + \pi_d (\delta_k - \bar{\delta}) \right. \\ \left. - \pi_{r,d} (\bar{e}_{r,k} - \bar{e}_r) - \pi_{r,d} (\bar{\mathbf{P}}_{r,d} \mathbf{b}_{r,o,k} - \bar{\mathbf{P}}_{r,d} \bar{\mathbf{b}}_{r,o}) - \pi_{r,d} (\delta_k - \bar{\delta}) \right]^2.$$

This has expectation,

$$E \left[\hat{B}_\delta \right] = \pi_d^2 \left[\frac{\sigma^2 + n_d \bar{\mathbf{P}}_{a,d} \mathbf{V}_{a,o} \bar{\mathbf{P}}_{a,d}^T}{n_d} \right] + \pi_{r,d}^2 \left[\frac{\sigma^2 + n_{r,d} \bar{\mathbf{P}}_{r,d} \mathbf{V}_{r,o} \bar{\mathbf{P}}_{r,d}^T}{n_{r,d}} \right] + \left(\frac{n_d - n_{r,d}}{n} \right)^2 \sigma_\delta^2.$$

This can also be expressed as,

$$E \left[\hat{B}_\delta \right] = E \left[\hat{B}_{\text{DJ}} \right] + \left(\frac{n_d - n_{r,d}}{n} \right)^2 \sigma_\delta^2.$$

The information anchored variance is,

$$E [V_{\text{anchored}}] = \mathbf{a}^T \mathbf{D}_{\text{DJ}} \boldsymbol{\Sigma} \mathbf{D}_{\text{DJ}}^T \mathbf{a} + O \left(\frac{\pi_d(1 - \pi_d)}{n} \right) + O \left(\frac{\pi_{r,d}(1 - \pi_{r,d})}{n} \right) \\ + E \left[\hat{B}_{\text{DJ}} \right] + \frac{E \left[\hat{B}_{\text{DJ}} \right]}{E \left[\hat{W}_{\text{DJ}} \right]} \left[O \left(\frac{\pi_d(1 - \pi_d)}{n} \right) + O \left(\frac{\pi_{r,d}(1 - \pi_{r,d})}{n} \right) \right].$$

If Rubin's variance estimator preserves the information loss seen in the primary analysis under MAR then,

$$E \left[\hat{W}_\delta \right] + E \left[\hat{B}_\delta \right] \approx \mathbf{a}^T \mathbf{D}_{\text{DJ}} \boldsymbol{\Sigma} \mathbf{D}_{\text{DJ}}^T \mathbf{a} + O \left(\frac{\pi_d(1 - \pi_d)}{n} \right) + O \left(\frac{\pi_{r,d}(1 - \pi_{r,d})}{n} \right) + E \left[\hat{B}_{\text{DJ}} \right] \\ + \frac{E \left[\hat{B}_{\text{DJ}} \right]}{E \left[\hat{W}_{\text{DJ}} \right]} \left[O \left(\frac{\pi_d(1 - \pi_d)}{n} \right) + O \left(\frac{\pi_{r,d}(1 - \pi_{r,d})}{n} \right) \right].$$

After simplification this is,

$$0 \approx E \left[\hat{B}_{\text{DJ}} \right] - E \left[\hat{B}_\delta \right] + \frac{E \left[\hat{B}_{\text{DJ}} \right]}{E \left[\hat{W}_{\text{DJ}} \right]} \left[O \left(\frac{\pi_d(1 - \pi_d)}{n} \right) + O \left(\frac{\pi_{r,d}(1 - \pi_{r,d})}{n} \right) \right].$$

Which is,

$$0 \approx - \left(\frac{n_d - n_{r,d}}{n} \right)^2 \sigma_\delta^2 + \frac{E \left[\hat{B}_{DJ} \right]}{E \left[\hat{W}_{DJ} \right]} \left[O \left(\frac{\pi_d(1 - \pi_d)}{n} \right) + O \left(\frac{\pi_{r,d}(1 - \pi_{r,d})}{n} \right) \right]. \quad (5.12)$$

With deviation in both arms, the difference between Rubin's variance estimator and the information anchored variance remains dominated by the variance of δ . In comparison to the result for the same setting, with deviation in the active arm only (5.8) we see there is an additional component in the difference, which depends on the variance of δ for the reference deviators. However, if there is a similar number of active deviators and reference deviators, i.e. $n_d \approx n_{r,d}$ there will be a closer approximation of the information anchoring variance by Rubin's variance estimator. In this case the largest term of the difference, which is also driven by σ_δ^2 (first term on the RHS of (5.12)) will be zero. The information anchoring is driven by how similar the proportion of deviation is in each arm and the variance of σ_δ^2 .

The second term on the RHS of (5.12) will remain negligible in practice, since as discussed previously $E \left[\hat{B}_{DJ} \right]$ is typically much smaller than $E \left[\hat{W}_{DJ} \right]$. Further the $E \left[\hat{B}_{DJ} \right] / E \left[\hat{W}_{DJ} \right]$ term is multiplied by other terms with an order that is less than n^{-1} .

We now consider a different δ -adjustment in each arm. That is we adjust by $\delta_{z,k} \sim N(\delta_z, \sigma_{\delta_z}^2)$ for patients in arm $z = a, r$. We assume the covariance between δ_a and δ_r is zero. When a different δ -adjustment is made in each arm,

$$\begin{aligned} \hat{B}_\delta = \frac{1}{K-1} \sum_{k=1}^K [& \pi_d (\bar{e}_k - \bar{e}) + \pi_d (\bar{\mathbf{P}}_{a,d} \mathbf{b}_{a,o,k} - \bar{\mathbf{P}}_{a,d} \bar{\mathbf{b}}_{a,o}) + \pi_d (\delta_{a,k} - \bar{\delta}_a) \\ & - \pi_{r,d} (\bar{e}_{r,k} - \bar{e}_r) - \pi_{r,d} (\bar{\mathbf{P}}_{r,d} \mathbf{b}_{r,o,k} - \bar{\mathbf{P}}_{r,d} \bar{\mathbf{b}}_{r,o}) - \pi_{r,d} (\delta_{r,k} - \bar{\delta}_r)]^2, \end{aligned}$$

where $\bar{\delta}_z = \frac{1}{K} \sum_{k=1}^K \delta_{z,k}$ for $z = a, r$. This has expectation,

$$\begin{aligned} E \left[\hat{B}_\delta \right] = & \pi_d^2 \left[\frac{\sigma^2 + n_d \bar{\mathbf{P}}_{a,d} \mathbf{V}_{a,o} \bar{\mathbf{P}}_{a,d}^T}{n_d} \right] + \pi_{r,d}^2 \left[\frac{\sigma^2 + n_{r,d} \bar{\mathbf{P}}_{r,d} \mathbf{V}_{r,o} \bar{\mathbf{P}}_{r,d}^T}{n_{r,d}} \right] \\ & + \pi_d^2 \sigma_{\delta,a}^2 + \pi_{r,d}^2 \sigma_{\delta,r}^2. \end{aligned}$$

This can also be expressed as,

$$E \left[\hat{B}_\delta \right] = E \left[\hat{B}_{DJ} \right] + \pi_d^2 \sigma_{\delta,a}^2 + \pi_{r,d}^2 \sigma_{\delta,r}^2.$$

For information preservation,

$$0 \approx -\pi_d^2 \sigma_{\delta,a}^2 - \pi_{r,d}^2 \sigma_{\delta,r}^2 + \frac{E[\hat{B}_{DJ}]}{E[\hat{W}_{DJ}]} \left[O\left(\frac{\pi_d(1-\pi_d)}{n}\right) + O\left(\frac{\pi_{r,d}(1-\pi_{r,d})}{n}\right) \right].$$

In this case the sharpness of the approximation will depend on the different variance's of δ in both arms. If the covariance between δ_a and δ_r is not zero then there will be an additional term in the difference between Rubin's variance estimator and the information anchored variance. This will represent the covariance between δ_a and δ_r . Thus the information anchoring performance of Rubin's variance estimator will also depend on the covariance between δ_a and δ_r .

We can similarly follow Proposition 5 through with the above amendment for differential δ -adjustment in each arm to show that the information anchoring performance of Rubin's variance estimator depends on the assumed variance of δ in both arms and the covariance between δ_a and δ_r .

It follows that if δ is fixed, Rubin's variance estimator will be excellent at information anchoring in the longitudinal trial setting with deviation in both arms.

Throughout this chapter we have assumed separate imputation models by treatment arm. If this is not the case then naturally not all of the imputation parameters for patients in arm a will be independent of the imputation parameters for patients in arm r , for both de-jure and δ imputation. Rubin's variance estimator will however still be information anchoring up to the same order. The same additional terms, representing the covariance between the imputation parameters of the different arms, will be incorporated in the expected value of the de-jure and δ between imputation variance. When looking at the difference between these quantities, the additional covariance components will cancel out and the results presented above will thus still follow.

As in Chapter 4, for simplicity we have also focussed attention on the setting with infinite imputations, that is for $K \rightarrow \infty$. Following the discussions in Section 4.6 we conclude that Proposition 4 and 5 also hold with finite K . Only smaller terms, additionally multiplied by $(1/K)$, will be incorporated in the approximation.

5.9 Summary

The ' δ -method' is widely used by trialists [16] (see for example [61, 62]). It is accessible to all members of the trial team and others with a stake in the trial since assumptions are framed with respect to how the data for the deviators differs to those observed. It can also be used for tipping point analysis, to establish under what circumstances alternative inference would be obtained.

Rubin's treatment estimator is unbiased and Rubin's variance estimator provides an excellent approximation for the gold standard information anchored variance we desire in the δ -adjusted MI sensitivity analysis setting. That is the design based variance we would observe if all δ -adjusted deviation data were present, adjusted for the loss of information seen in the primary MAR analysis.

The approximation is better when δ has a small variance, or indeed none and is a fixed quantity. A large variance on δ reflects uncertainty on the difference in response between observed cases and deviators, hence it is unsurprising that the approximation is better when we are more certain of the difference. When δ is not fixed Rubin’s variance estimator loses information relative to the information anchored variance i.e. it will overestimate the information anchored variance. When δ varies by treatment arm and is not fixed, the information anchoring performance of Rubin’s variance estimator will further depend on the covariance of the δ ’s in the different arms.

This highlights the importance of the elicitation of the sensitivity analysis parameters. This is notoriously difficult, as described in [28]. In some cases trialists may be confident of a set of fixed δ adjustments, which correspond to credible scenarios for the unobserved patients, to use in sensitivity analysis. When exact values of δ are uncertain, to be information anchoring in sensitivity analysis a greater number of fixed δ adjustments should be made over plausible values of δ . If a prior for δ with a large variance on δ is alternatively used the variance of the resulting treatment estimator will not be information anchored. Exploring the uncertainty of δ via a greater range of fixed δ will ensure the information lost due to missing data in the primary analysis is preserved in the sensitivity analysis, with no additional loss of information.

Often in practice, in a single trial setting, a group of clinical experts can be approached to provide values of δ to use in sensitivity analysis. That is the differences in the expected mean response between those observed and those unobserved. If one or more clinician provides a value of δ that is considerably higher or lower than the others this could result in large variance in δ over the experts. Incorporating this information into a single prior for δ will then lead to a loss of information in the sensitivity analysis. If alternatively one runs a number of fixed δ -adjusted analyses, covering the range of proposed δ ’s, an information anchored variance estimate will be obtained in each case.

The fact that the information anchoring behaviour of Rubin’s variance estimator is seen in both the δ -adjusted and reference based MI settings only further supports the use of Rubin’s variance estimator in the reference based MI settings. Whilst Rubin’s variance estimator does not match the long-run sampling variance of the reference based MI estimator —appropriately as the long-run sampling variance is unnatural and artificially small due to borrowing data from within the trial for sensitivity analysis— it does have the same information anchoring behaviour as that observed in the δ setting. In the δ setting when the long-run sampling variance is not unnatural, the information anchored variance approximates the long-run sampling variance.

Chapter 6

Software and application

Despite the increasing volume of research on more sophisticated statistical methods for dealing with missing data and sensitivity analysis in high impact journals [1, 63, 64], a recent review by Bell et al. found relatively little of these methods in practice [5]. Out of 77 RCTs published in four top leading medical journals in 2013, 73 (95%) reported missing outcome data and the most commonly used method in the primary analysis was complete case analysis 33 (45%). Simple imputation was performed by 20 (27%) trials, 15 (19%) used model based methods and only 6 (8%) used MI. Further only 25% of the trials with missing data conducted sensitivity analysis that altered the missing data assumption made in the primary analysis. Many reasons can be postulated for this. As discussed by Bell et al.

“Perhaps the reluctance to use more sophisticated approaches is due to a lack of knowledge or experience on the parts of applied researchers and/or biostatisticians. Perhaps it is due to the time lag between reports of methods and software to implement them.”[5]

A cross-sectional survey of members of the Statistical Society of Canada (SSC) in 2014 investigating barriers to the uptake of new statistical methods also identified lack of suitable software as a key factor (selected by 81% statisticians) [65]. We agree that the availability of user friendly software to implement novel techniques along with practical examples demonstrating their application is fundamental.

Although the use of more developed statistical techniques for handling missing data, such as MI, still falls behind simpler methods of analysis, a review of RCTs published in four leading medical journals by Mackinnon, from the earliest date each journal offered searches until the end of 2008, shows MI has gained popularity over the years [66]. The review highlights a substantial increase in use from 2005 to 2008. A more recent review of publications in two leading medical journals from 2008 to 2013 by Rezvan et al. [67], identified a continuing increasing trend in the number of RCTs using MI over time.

This positive trend is encouraging and highlights the impact of accessible software, since it coincides with the introduction of implementations of MI in mainstream statistical software packages, such as the MICE implementation in Stata by Royston in 2004 [68]. This was followed by the introduction of a suite of inbuilt MI commands in Stata release 11 in 2009 [69], which was updated in Stata

release 12 in 2011 to incorporate the MICE implementation [70]. SAS introduced MI in version 8.2 in 2001 [71] using the MVN model, which was updated in version 9.3 in 2011 [72] to include the MICE approach. MICE was also introduced in SPSS version 17.0 in 2008 [73]. Various implementations of MI were also introduced in R during this time period.

In this thesis we have seen MI is a valuable technique for trial analysis not only under the MAR assumption but in sensitivity analysis, under MNAR assumptions. However of the 103 papers reviewed by Rezvan [67] between 2008 and 2013 that used MI, only 3 (3%) performed a sensitivity analysis following MI to investigate the robustness of the MI estimate to departures from the MAR assumption. Of these only one paper provided details about the sensitivity analysis approach. The review highlights a “lack of explicit guidelines in the literature for conducting sensitivity analyses within the MI framework.”

Carpenter, Roger and Kenward (2013) address this need with the introduction of the reference based MI procedures evaluated in this thesis [13]. The procedures were introduced with a SAS macro written by Roger, freely available at www.missingdata.org for implementation [74]. Aiming to increase the uptake of their approach a user friendly Stata programme named `mimix` has been created [56]. Stata is a statistical software package, used globally by statisticians in the medical and other industries.

In the next section we introduce the new Stata command `mimix`. In Section 6.2 we demonstrate the use of `mimix`, using data from the asthma trial. These two sections form part of a manuscript published in the Stata journal by myself and my supervisors [55]. In Section 6.3 we present a second application and analyse the reviewer study data. Stata code is presented for all the analyses to facilitate adoption. We finish with a discussion and presentation of the download statistics for `mimix`, up to the time of writing of this thesis in Section 6.4.

6.1 Introducing `mimix`

The `mimix` command conducts reference based MI under the five distinct treatment arm based assumptions for missing data outlined in Section 2.1 (MAR, CR, J2R, CIR and LMCF). The user can specify the required imputation method for all individuals, or is able to specify individual-specific imputation methods. Additionally an alternative imputation method can be specified for all interim missing values; where the individual has later observed data. Often it may be desirable to impute interim missing data under randomised-arm MAR. Or the interim missing data can be treated in the same way as missing post-deviation data.

Optionally, two substantive analysis models can be fitted to each imputed dataset, and the results summarised using Rubin’s rules, by `mimix`. The two analysis options in `mimix` are (a) a linear regression of the outcome at the final time point on treatment and baseline, or (b) a saturated repeated measures model (i.e. including treatment crossed with visit and baseline crossed with visit), with separate covariance matrices for each treatment arm. Other substantive analysis models can be fitted to the imputed data in the usual way using the usual Stata `mi estimate` commands post-completion of `mimix`.

`mimix` is freely available for download at,

<https://ideas.repec.org/c/boc/bocode/s457983.html> [56].

6.1.1 Syntax

The syntax of the `mimix` command is the following:

```
mimix depvar treatvar, id(varname) time(varname) [ clear covariates(varlist)
interim(string) iref(string) method(string) methodvar(varname) mixed refgroup(string)
refgroupvar(varname) regress saving(filename[,replace]) burnbetween(#) burnin(#)
m(#) seed(#) ]
```

Data are required in long format with one record per individual per time point, where *depvar* is the numeric outcome variable with missing data in the existing dataset and *treatvar* identifies the treatment group variable in the existing dataset and may be either a numeric or string variable.

6.1.2 Options

Here we describe the options available when using the command.

`id(varname)` specifies the variable identifying individuals in the existing dataset. This is a required option and may be either a numeric or string variable.

`time(varname)` specifies the variable identifying units of time in the original dataset. This is a required option and must be a numeric variable.

`clear` specifies that the original data in memory be cleared and replaced by the imputed dataset. The imputed dataset must then be saved manually if required. `clear` and/or `saving()` is required.

`covariates(varlist)` specifies any additional baseline covariates to be included in the imputation model, and analysis if either the `regress` or `mixed` options are specified. Any specified covariates must be fully observed numerical variables. Dummy variables must be generated for any factor covariates.

`interim(string)` specifies an alternative imputation method for all interim missing values; where the individual has later data observed. Values must be one of “mar”, “j2r”, “lmcf”, “cir” or “cr” (not case sensitive). See Subsection 6.1.4 for further details on specifying the imputation method.

`iref(string)` specifies the level of *treatvar* chosen for the reference for all interim missing values; where the individual has later data observed. Required when using the “j2r”, “cir” and “cr” imputation methods. See Subsection 6.1.4 for further details on specifying the imputation method.

`method(string)` defines the imputation method for all individuals. Values must be one of “mar”, “j2r”, “lmcf”, “cir” or “cr” (not case sensitive). See Subsection 6.1.4 for further details on specifying the imputation method.

`methodvar(varname)` specifies the variable in the original dataset containing individual specific imputation method(s). This option should be used if different imputation methods are required for different individuals. `methodvar()` must be a string variable containing values of “mar”, “j2r”, “lmcf”, “cir” or “cr” (not case sensitive). `methodvar()` and `method()` are mutually exclusive; specifying both will return an error message. See Subsection 6.1.4 for further details on specifying the imputation method.

`mixed` uses `mi estimate` with Stata’s default options to fit a saturated repeated measures model using restricted maximum likelihood—with a separate mean for each treatment and time together with full covariate–time interactions for any included `covariates()` and a separate unstructured covariance matrix for each arm—to each of the imputed datasets and combines results using Rubin’s rules for inference. It should be noted that this option may add substantially to the post-imputation computation time if a large number of imputations have been specified.

`refgroup(string)` specifies the level of *treatvar* chosen for the reference for all individuals. Required when using the “j2r”, “cir” and “cr” imputation methods. See Subsection 6.1.4 for further details on specifying the imputation method.

`refgroupvar(varname)` specifies the variable in the original dataset which identifies the level of *treatvar* chosen for the reference for each individual. Required when using the “j2r”, “cir” and “cr” imputation methods. `refgroupvar()` and `refgroup()` are mutually exclusive; specifying both will return an error. See Subsection 6.1.4 for further details on specifying the imputation method.

`regress` uses `mi estimate` with Stata’s default options to fit a linear regression of *devar* at the final time point on *treatvar*, and any included `covariates()`, to each of the imputed datasets and combines results using Rubin’s rules for inference.

`saving(filename [,replace])` saves the imputed datasets. A new file name is required unless `replace` is also specified. `replace` allows the *filename* to be overwritten with new data and may not be abbreviated. `clear` and/or `saving()` is required.

`burnbetween(#)` specifies the number of iterations between pulls for the posterior in the MCMC. The default is 100.

`burnin(#)` specifies the number of iterations in the MCMC burn in. The default is 100.

`m(#)` specifies the number of imputations required. The default is 5.

`seed(#)` specifies the seed for the random number generator. The default is 0 meaning that no seed is specified by the user and the current value of Stata’s random number seed will be used; this will consequently result in different sets of imputations for multiple programme runs. To reproduce a set of imputations the same random number seed should be used with the original data sorted in exactly the same order.

6.1.3 Implementation details

Baseline covariates

Any additional baseline covariates that are included are required to be complete. Individuals with missing covariate information will be highlighted to the user by the command and discarded in the imputation process and any requested analysis.

Potential error with sparse data

Stata's `mi impute mvn` command, which uses the MCMC method initialised by the EM algorithm to impute missing values, is utilised to complete steps 1 and 2 of the general procedure, as detailed in Section 2.1. If the response variable of interest is measured at an occasion with only a very few complete cases `mi impute mvn` may terminate with an error if there is not enough information in the observed data to reliably estimate aspects of the covariance structure in the required MVN model. If this is the case we advise the user to explore an alternative viable MVN model for the data using the `mi impute mvn` command. The response at the occasion with few observed outcomes may potentially need to be excluded from the analysis and `mimix` re-run.

Data output

The imputed datasets are produced in long format, with one record per individual per time point per imputation, and are `mi set` in Stata's `flong` style, ready to analyse using `mi estimate` in Stata. The imputed datasets are output in memory if `clear` is specified and/or saved in *filename.dta* if `saving()` is specified.

Analysis options

If the `regress` or `mixed` analysis options are specified Stata's `mi estimate` command is utilised to fit the specified analysis model to each imputed dataset and combine results using Rubin's rules. The usual MI output will be displayed in the results window. If other substantive analysis models are required following the completion of `mimix`, `mi estimate` can be used in the usual way for further analysis.

Use of data preserve

Due to extensive manipulation of the data `mimix` uses Stata's `preserve` and `restore` commands. Whilst `mimix` can be successfully run on data that are already preserved, we recommend users cancel any previous data preserve using `restore`, not to ensure the `clear` and `saving()` options of `mimix` work as intended.

6.1.4 Specifying the imputation method

Either the `method()` option, to indicate which imputation method should be employed for all individuals, or the `methodvar()` option, to indicate which imputation method should be employed for each individual must be used. `method()` and `methodvar()` are mutually exclusive options; specifying both will return an error.

If the `method()` option is used to request the same imputation method for all individuals, then values specified in `method()` must be one of those presented in Table 6.1 (not case sensitive). If the `methodvar()` option is used to request different imputation methods for different individuals, then a new variable that contains individual specific imputation methods must be generated and specified in `methodvar()`. The variable that holds the individual imputation methods must only contain values presented in Table 6.1 (not case sensitive) and the method specification cannot vary within an individual over time.

If either of the J2R, CIR or CR options are used then either the `refgroup()` option must also be used to specify the reference level of the *treatvar* for all individuals, or the `refgroupvar()` option to indicate the reference level of the *treatvar* for each individual. Together these variables allow for the required assumptions outlined in Section 2.1 [13]. If one of the imputation methods that includes a reference group is specified for all individuals (or for specific individuals via `methodvar()`) missing data for individuals in that reference group (with the reference imputation specification) are imputed under *randomised-arm MAR*.

The `interim` option specifies the imputation method for all interim missing values, that is where individuals have later data observed. If this option is not selected any interim missing values will be imputed following the method specified by the `methodvar()` or `method()` option, in the same way as missing post-deviation data.

Method name	Name to specify in <code>method()</code> or <code>methodvar()</code>
Randomised-arm Missing At Random	mar
Jump to reference	j2r
Last mean carried forward	lmcf
Copy increments in reference	cir or ciir
Copy reference	cr

Table 6.1: Specifying the imputation method using `mimix`. Names to specify are not case sensitive.

6.2 Sensitivity analysis of the asthma trial

We now conduct reference based sensitivity analysis using the `mimix` command for the randomised double-blind placebo controlled trial of budesonide delivered by Turbuhaler for the treatment of adult patients with chronic asthma described in Subsection 1.7.2. Analysis was performed using Stata version 14.1 [75].

The observed mean FEV₁ profiles by treatment arm (active vs. placebo) and the various missing data patterns are shown in Figure 1.3. Only 38 of the 92 individuals in the placebo arm (41%) and 72 of the 91 individuals in the active arm (79%) remained in the trial at 12 weeks; 3 individuals (2 placebo and 1 active) had interim missing data.

The primary analysis of the original trial consisted of a linear regression of the 12 week FEV₁ outcome on treatment group, adjusted for baseline FEV₁, using data from the 110 individuals measured at week 12 (complete case analysis under MAR). This gives a treatment effect of 0.239 L, $p = 0.017$. We will use `mimix` to assess the robustness of the results to the various post-deviation assumptions outlined in Section 2.1. The interim missing outcomes will be imputed under MAR.

In the following output we describe the variables in the asthma trial dataset and list their contents for one arbitrarily selected deviating individual.

```
. use asthma, clear
. describe
Contains data from asthma.dta
  obs:          732
  vars:          5                      12 Feb 2015 10:18
  size:         11,712
```

variable name	storage type	display format	value label	variable label
id	int	%8.0g		Patient ID
time	byte	%9.0g		Measurement time (weeks)
treat	byte	%8.0g	treat1	Randomised treatment assignment
base	double	%12.0g		Baseline FEV1 (L)
fev	float	%9.0g		FEV1 (L)

```
Sorted by: id
. list in 37/40, noobs sepby(id)
```

id	time	treat	base	fev
5030	2	Placebo	1.14	.85
5030	4	Placebo	1.14	1.51
5030	8	Placebo	1.14	.
5030	12	Placebo	1.14	.

`id` is the unique individual identifier and `treat` is the randomised treatment assignment, to placebo (`treat = 2`) or active (`treat = 3`). `fev` is the post-baseline FEV₁ measurement (recorded in Litres) and `time` is the time of the FEV₁ measurement in weeks. `base` is the baseline FEV₁ measurement. We can see that the selected individual deviated sometime between week 4 and week 8; consequently, the individual has missing outcomes for weeks 8 and 12. We also note that the dataset is already in long format with one observation per individual per time point, as required for `mimix`.

We first analyse the data under a more realistic MAR assumption using previously observed outcomes for all individuals; in other words a de-jure assumption that—post-deviation—individuals continued on their randomised treatment as specified in the protocol. We create 50 imputations and take the default MCMC burn-in of 100 and burn-between of 100 iterations. We include the baseline FEV₁ measure in the imputation model as a covariate, but if this fully observed variable

were used as an outcome, the results would be stochastically identical. We use the `regress` option to specify that the substantive analysis is a linear regression of 12-week FEV₁ on randomised treatment and baseline FEV₁. Imputation with the MAR option automatically means the interim missing values will be imputed under MAR in each treatment group. The command required and produced results are shown in the following output.

```
. mimix fev treat, id(id) time(time) method(mar) covariates(base) regress m(50)
> clear seed(101)
Performing imputation procedure for group 1 of 2...
Performing imputation procedure for group 2 of 2...

Performing regress procedure ...

i.treat      _Itreat_2-3      (naturally coded; _Itreat_2 omitted)

Multiple-imputation estimates      Imputations      =      50
Linear regression      Number of obs      =      183
                        Average RVI      =      0.4106
                        Largest FMI      =      0.3495
                        Complete DF      =      180
DF adjustment:      Small sample      DF:      min      =      91.39
                        avg      =      99.15
                        max      =      105.79
Model F test:      Equal FMI      F( 2, 149.8)      =      40.69
Within VCE type:      OLS      Prob > F      =      0.0000
```

fev	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
_Itreat_3	.3230728	.1042794	3.10	0.002	.1163241 .5298215
base	.7240691	.0861441	8.41	0.000	.5531672 .8949709
_cons	.3959986	.1971734	2.01	0.048	.0043602 .787637

```
Imputed dataset now loaded in memory
Imputed data created in variable fev using mar
```

The output displays the results from the MI linear regression analysis requested, along with a description of the variable which now contains imputed data. Under MAR MI, using the previously observed outcomes, the treatment estimate is increased from the complete case regression reported above, to 0.323 L, with a p-value of 0.002. The results of this analysis are shown in the top panel of Figure 6.1. The complete case analysis makes a stronger assumption that the conditional distributions of missing and observed 12 week responses given treatment and baseline are the same. Whilst the MI MAR analysis assumes the conditional distributions of missing and observed 12 week responses given treatment, baseline and previous observed responses are the same, which is more likely to be realistic.

As the `clear` option was specified, the imputed dataset is stored in memory. The imputed data are output `mi set` in `flong` format, using `mi set flong`. We note that the imputed dataset has not yet been saved. If the `saving()` option is specified when using `mimix` then the imputed data will be saved when the command is executed.

We now re-impute the asthma trial, under the J2R assumption for all individuals, with the placebo arm (`treat =2`) first set as the reference. The `interim` option is included to impute the interim missing values under MAR. We note that inclusion of the `interim` option here does not actually effect the results since our substantive model of interest considers the treatment effect at the final time point. Imputation of interim values under MAR (or another reference based approach) will have an impact when the `mixed` option is specified to fit a saturated repeated measures model,

using all follow-up outcomes, to estimate a separate baseline adjusted treatment effect at each follow-up time.

```
. mimix fev treat, id(id) time(time) method(j2r) refgroup(2) covariates(base) i
> nterim(mar) regress m(50) clear seed(101)
Performing imputation procedure for group 1 of 2...
Performing imputation procedure for group 2 of 2...
```

Performing regress procedure ...

```
i.treat      _Itreat_2-3      (naturally coded; _Itreat_2 omitted)

Multiple-imputation estimates      Imputations      =      50
Linear regression                  Number of obs    =     183
                                   Average RVI      =     0.4483
                                   Largest FMI     =     0.3510
                                   Complete DF    =     180

DF adjustment:  Small sample      DF:      min    =     91.07
                                   avg        =    109.09
                                   max        =    140.18

Model F test:      Equal FMI      F(  2, 156.9) =     32.45
Within VCE type:  OLS            Prob > F      =     0.0000
```

fev	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
_Itreat_3	.2261827	.1028346	2.20	0.029	.0228754	.42949
base	.6894261	.0933944	7.38	0.000	.5040403	.8748119
_cons	.4669997	.2112431	2.21	0.030	.0473954	.8866041

```
Imputed dataset now loaded in memory
Imputed data created in variable fev using j2r
Interim missing data imputed using mar
```

The results of the J2R analysis with placebo as the reference are summarised in Table 6.2 along with results of a J2R analysis with active as the reference. These address the de-facto assumption, when post-deviation individuals not randomised to the reference treatment change to the reference treatment. Both these analyses result in a reduced treatment estimate relative to the de-jure MAR assumption. However, while J2R with placebo as the reference still gives a treatment effect that is statistically significant at the 5% level, J2R with active as the reference does not. This is because more individuals deviate in the placebo arm than the active arm (Figure 1.3) and they tend to be individuals whose lung function is lower. The effect of this versus analysis under MAR is shown in Figure 6.1. The change in placebo individuals under J2R-active reduces the treatment estimate by the greatest amount.

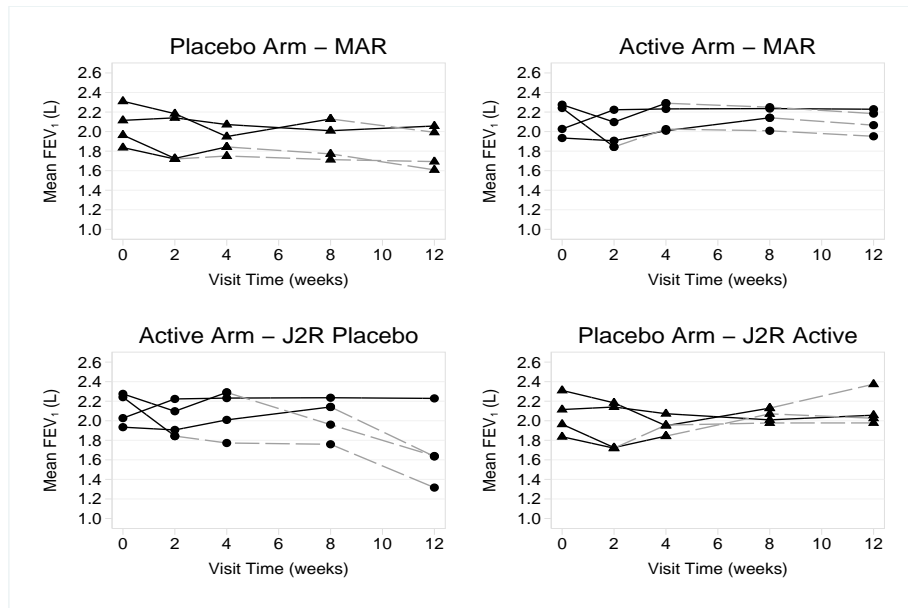


Figure 6.1: Mean FEV₁ against time, by treatment arm, for the four different deviation (withdrawal) patterns under MAR (top panel) and J2R (bottom panel). Solid lines join observed means before deviation and dashed lines join the means of the imputed data for that pattern.

Our next analysis is LMCF. Figure 1.3 shows that the arm-specific means begin to stabilise quite early in the follow-up. It is therefore to be expected that LMCF gives a slightly reduced estimate relative to MAR, with a slightly higher p-value (Table 6.2). If we wish to assume that individuals' lung function at deviation is broadly maintained post-deviation, LMCF would be appropriate.

The next two analyses are both CIR. In the first, the reference is the placebo, and in the second, the reference is the active arm. Because more individuals deviate in the placebo arm than the active arm, and the placebo arm profiles tend to decrease while those of the active arm increase, we again see a slightly larger treatment estimate when the reference arm is placebo (Table 6.2). CIR—when the reference is placebo—is appropriate if, post-deviation, we wish to assume that active individuals' lung function starts to decline from its current value at the same rate as seen in the placebo arm. The converse—when the reference is active—is appropriate if, post-deviation, we wish to assume that placebo individuals access an active treatment and their lung function increases from its current value at the rate seen in the active arm.

Finally, we consider CR, with placebo and active reference. Under this assumption, an individual's post-deviation data are imputed as if they had always belonged to the reference arm (in effect, received the reference arm treatment). CR with placebo reference may be an appropriate de-facto assumption for individuals who could not tolerate the active treatment. Under CR, pre-deviation individual-specific residuals about the mean are typically greater than under J2R. This means that post-deviation profiles typically change less abruptly than with J2R. This is what we observe here (Table 6.2). For CR with active reference, the treatment estimate is greater than both treatment estimates under J2R, but less than the treatment estimates for all the other de-facto assumptions.

Analysis	Treatment Est. (L)	Std. Err.	<i>p</i> -value
De-jure			
Primary Analysis (ANCOVA)	0.239	0.099	0.017
Randomised-arm MAR	0.323	0.104	0.002
De-facto			
Jump to placebo	0.226	0.103	0.029
Jump to active	0.128	0.095	0.181
Last mean carried forward	0.296	0.096	0.003
Copy increments in placebo	0.281	0.103	0.007
Copy increments in active	0.277	0.082	0.001
Copy placebo	0.289	0.101	0.005
Copy active	0.251	0.082	0.003

Table 6.2: Sensitivity analysis results for the asthma trial. $K = 50$ imputations.

We therefore conclude that if post-deviation medication has a comparable effect to the lowest active dose, then individuals will have comparable lung function at the end of the study. Otherwise the sensitivity analysis is consistent with the primary analysis of the trial, in identifying a significant beneficial effect of treatment relative to placebo.

The results in Table 6.2 demonstrate the information anchoring behaviour of Rubin’s variance estimator in a real setting. We see the information loss in the MI MAR analysis is generally held reasonably constant over the sensitivity analyses, which also condition on previously observed outcomes. The standard errors of the treatment estimates in the de-facto imputation scenarios are not drastically different in any of the analyses. The standard error does tend to be marginally smaller for reference based methods where the active group is the reference arm. Only 38 of the 92 individuals in the placebo arm (41%) and 72 of the 91 individuals in the active arm (79%) remained in the trial at 12 weeks. Since there were more deviators in the placebo arm of this trial this small difference is not surprising given the theory in Chapter 4. There will be a greater proportion of reference based deviation when the active group is the reference arm. The approximation between Rubin’s variance estimate and the information anchored variance is not as sharp as the components which form the difference between Rubin’s variance and the information anchored variance will be slightly larger than when the reference group is the placebo arm.

6.3 Sensitivity analysis of the reviewer study

We now conduct sensitivity analysis for the reviewer study described in Section 1.7.1. The primary analysis of the original trial, performed under the MAR assumption (consisting of a linear regression of review quality on treatment group, adjusted for baseline) gave a statistically significant treatment effect of 0.237 improvement in quality, 95% CI [0.099-0.376], $p = 0.001$. However 46 participants dropped out the self-taught group compared with only 11 in the no training group. Moreover those who dropped out in the self-taught group had a significantly poorer review quality at baseline. It is therefore important to assess how robustness the conclusions are to the MAR assumption.

Carpenter and Kenward [44] have previously conducted sensitivity analysis of the reviewer data using the ‘ δ -method’. A questionnaire was developed to elicit experts (20 editors and other staff at the BMJ) prior opinion about the average difference in review quality index between those who were observed and those who were not [28]. The resulting distribution was $N(-0.21, 0.46^2)$. We first replicate the δ -adjusted sensitivity analysis performed by Carpenter and Kenward assuming $\delta \sim N(-0.21, 0.46^2)$. We then conduct a further δ -adjusted sensitivity analysis which assumes δ is fixed at -0.21 i.e. we ignore the variability across the experts and take into account just the average anticipated difference in review quality index by completer status. We then establish what the results would look like if we assume the reviewers who did not return paper 2 behaved like those in the no training group.

In the following output we describe the variables in the reviewer dataset and list their contents for a randomly selected individual.

```
. use reviewer, clear
. describe
Contains data from reviewer.dta
  obs:      339
  vars:      4                               16 Jan 2009 15:57
  size:     5,424
```

variable name	storage type	display format	value label	variable label
id	float	%9.0g		Reviewer identifier
inter	float	%19.0g	intervention	Training package
base	float	%9.0g		Paper 1 (baseline) mean review quality
resp	float	%9.0g		Paper 2 (response) mean review quality

```
Sorted by:
. list in 6, noobs sepby(id)
```

id	inter	base	resp
25	no training	1.714286	2.928571

`id` is the unique reviewer identification number, `inter` indicates the randomly assigned intervention package of no training (`inter`= 0) or self-taught package (`inter`= 1), `base` is the baseline mean review quality and `resp` is the mean review quality response for paper 2. We conduct the original primary analysis of the trial, a standard MI analysis under MAR and the original δ -adjusted sensitivity analysis using the below code. For each of the imputation analyses, 50 imputations were run. The analysis model was a linear regression of review quality on treatment group, adjusted for baseline as in the original primary analysis.

```
. regress resp base inter
. mi set flong
. mi register impute resp
. mi impute mvn resp=base, add(50) rseed(23) by(inter)
. mi estimate: regress resp base inter
. set seed 32
. bysort _mi_m: gen delta_k=(-0.21 + 0.46*invnormal(uniform())) if _n==1
```

```

· bysort _mi_m (delta_k):  replace delta_k=delta_k[1]
· mi passive:  gen imputed=_mi_miss
· replace imputed=1 if imputed==.
· gen resp_delta=resp+imputed*delta_k
· mi estimate:  regress resp_delta base inter

```

For the fixed δ -adjusted sensitivity analysis ($\delta \sim N(-0.21, 0)$) we replicated the above code with `delta_k=-0.21` in place of `delta_k=(-0.21 + 0.46*invnormal(uniform()))`. We use `mimix` to establish the treatment effect if the reviewers who did not return paper 2 copied the no training arm behaviour (`inter= 0`). The commands required are as follows.

```

· use reviewer, clear
· gen time = 2
· mimix resp inter, id(id) time(time) covariates(base) method(cr) refgroup(0)
  m(50) regress clear seed(23)

```

Results from analyses of the peer review study are summarised in Table 6.3. We see the intervention effect under copy no training behaviour is slightly reduced at 0.172, compared with 0.237 from complete case analysis and 0.234 from standard MI analysis (both under MAR), but remains statistically significant. We obtain an information anchored variance estimate in the copy control analysis with a standard error of 0.069 versus 0.071 under MAR MI and 0.070 from a complete case MAR analysis.

When taking into account the experts view, the effect of the self-taught intervention is no longer significant. The treatment effect is reduced to 0.195 and the standard error increased to 0.132. Allowing for the anticipated worse response from participants who dropped out, and the variability across the experts in their opinion on this, there is no evidence that the intervention improves the quality of peer review. The increased standard error represents the increased uncertainty about the effect of the dropout mechanism under the MNAR mechanism, relative to MAR. When this is taken into account the treatment effect is not statistically significant. We no longer obtain an information anchored variance estimate.

When we ignore the variability across the experts and just take into account the mean expert opinion we see the treatment effect is similarly decreased to 0.189. However, it remains significant with a standard error of 0.072. When δ is fixed we obtain an information anchored variance estimate in the sensitivity analysis. The standard error of the treatment estimate under fixed delta (0.072) is in close agreement with the standard of the treatment estimate under MAR (0.071 using MI and 0.070 from a complete case analysis).

These results corresponds with the theory presented in Chapters 4 and 5. In Chapter 5 we established that the information anchoring performance of Rubin's variance estimator in δ -adjusted sensitivity analysis is driven by the assumed variance for δ . This is why we no longer get an information anchored estimate when we take into account the variability across the experts and assume a variance of 0.46^2 for δ . This demonstrates the importance the assumed variance for δ has in sensitivity analysis.

Analysis	Treatment Est.	Std. Err.	<i>p</i> -value
Complete case (MAR)	0.237	0.070	0.001
MI, ‘on-treatment MAR’	0.234	0.071	0.001
MI, copy no training	0.172	0.069	0.013
MI, expert opinion $\delta \sim N(-0.21, 0.46^2)$	0.195	0.132	0.145
MI, mean expert opinion $\delta \sim N(-0.21, 0)$	0.189	0.072	0.009

Table 6.3: Sensitivity analysis results for the reviewer study. $K = 50$ imputations.

6.4 Discussion

We have demonstrated the information anchoring properties of Rubin’s variance estimate in two real life scenarios. The results correspond with the theory developed in this thesis for reference based sensitivity analysis and δ -adjusted sensitivity analysis. The results also highlight the impact of allowing for uncertainty in δ , in δ -adjusted sensitivity analysis. In the peer review study example when the variability across experts on the proposed difference between the completers and deviators was taken into account, the resulting variance estimate was no longer information anchored. The inference from the primary analysis was consequently completely overturned.

In practice trialists should consult with experts in their field to determine plausible values of δ a-priori for sensitivity analysis. Trialists must ensure they carefully consider the sensitivity parameters to investigate. If a number of experts are asked about the difference in outcome between those observed and those who deviate, in many cases most may be in consensus. But one or two experts with an extreme opinion may result in a large uncertainty surrounding the true value for δ . As we have seen, this can greatly effect whether the treatment estimator variance in sensitivity analysis is information anchored. Careful elicitation of sensitivity analysis parameters should be undertaken to ensure the most appropriate values are used. An advantage of reference based analysis is that no formal specification of sensitivity analysis parameters is required. When conducting reference based analysis using MI we also obtain an information anchored treatment estimator variance, as shown for the asthma and peer review RCTs.

The Stata program `mimix` successfully automates the reference based MI algorithm of Carpenter, Roger and Kenward [13] thus provides a computationally accessible tool for reference based sensitivity analysis. We hope that this implementation will remove a barrier to busy trialists performing sensitivity analysis in practice. The analysis of the asthma trial illustrates just how relevant and accessible the methodology is using `mimix`. The reference based sensitivity analysis of the reviewer study also demonstrates the utility of the approach. Stata code was presented for trialists to adopt and tailor to their own analysis needs.

Different qualitative assumptions can be made for different individuals, or similar groups of individuals, and the software allows this flexibility through the `methodvar()` option. This gives great flexibility for contextually plausible sensitivity analyses. Throughout this chapter we have focused

on the two-arm randomised trial setting, however this is not a constraint. `mimix` can be used to conduct reference arm-based imputation for trials with more than two arms.

Figure 6.2 illustrates downloads of `mimix` over time, from when it was made available for download in March 2015 until October 2016 (data beyond this not available at the time of writing). A total of 502 downloads have been recorded during this time period. On average `mimix` has received 26 downloads per month. A clear spike in downloads is evident in Figure 6.2 in June 2015. The increased demand corresponds with the teaching of an advanced missing data course at LSHTM where the `mimix` command has been successfully incorporated into a taught practical session. We are encouraged by these statistics which suggest `mimix` is currently making an impact and enabling relevant accessible sensitivity analysis of trials with missing data. We hope `mimix` will contribute to an increased use of the reference based sensitivity analysis methodology.

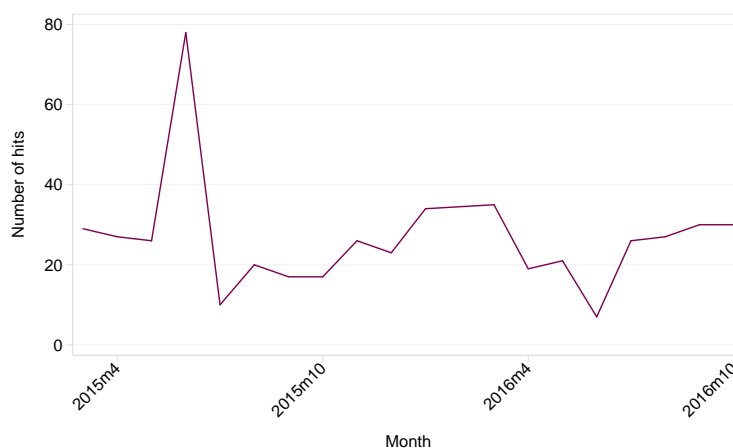


Figure 6.2: Mimix downloads

Chapter 7

Discussion

Undoubtedly the best way to handle missing data is to prevent its occurrence. However, realistically in clinical trials there will always be developments beyond the control of the investigators which result in missed patient visits or unrecoverable outcome measures. Primary analysis should be conducted under the most credible assumption for the missing data. Since we will never definitively know the exact mechanisms resulting in the missing data, it is important to perform sensitivity analysis under alternative plausible assumptions for the missing data. Regulatory guidelines stress the importance of this [11, 12], however they also highlight a need for accessible and relevant methods of sensitivity analysis.

In this thesis we have investigated and illustrated what have been termed controlled MI techniques, for sensitivity analysis of longitudinal RCT's with continuous missing data. Specifically the novel reference based MI procedures of Carpenter, Roger and Kenward [13] and the ' δ -method' presented in Carpenter and Kenward [7]. Both approaches retain the design based analysis model used in primary analysis in the sensitivity analysis, in order to assess purely the impact of alternative missing data assumptions on the primary outcome of interest. Rubin's rules are used to combine results across imputed data sets and provide one overall treatment estimate and an associated estimate of variance.

Rubin's variance estimator requires the imputation and analysis models to be congenial. That is the imputation and analysis model must have the same content and structure and thus be formed around the same assumptions. The assumptions of the design based analysis model however do not necessarily correspond with the assumptions made in imputation for both of the outlined controlled MI techniques. In each setting we therefore focussed on the evaluation of Rubin's variance estimator for the resulting treatment effect. This prompted us to first think about principles for variance estimation in sensitivity analysis of the type proposed, where the design based analysis model is retained.

7.1 Principles for sensitivity analysis

In sensitivity analysis we have to specify a distribution for the missing data. With continuous data this requires use to specify a mean and variance. It is possible that we could then have more information in the sensitivity analysis than we would have had, if we had actually observed the data. On the other extreme, it is quite possible to be so imprecise there is much less information in the sensitivity analysis than in the primary analysis. Conclusions from the primary analysis can then be overturned. Sensitivity analyses should not be injecting information ‘by the back door’ or taking it away. This is a concern of regulators and those in industry and other trial settings. This led to two key principles:

1. *Lower bound for variance estimation:* The expected value of the design based variance estimator if we were able to observe all the post-deviation data under the specified scenario provides a natural lower bound for the expectation of the variance estimator in sensitivity analysis.
2. *Information anchoring principle:* The information loss due to missing data in the primary analysis should equal the loss of information due to missing data in all sensitivity analyses.

That is firstly we cannot have more information in the sensitivity analysis than when the deviation data is actually observed under the postulated scenario. A loss of information is inevitable with missing data. We must account for this. Our estimate of variance must be scaled up to account for the missing information. Secondly a natural principle for the treatment estimator variance is to keep the information loss due to missing data constant across primary and all sensitivity analyses. Anchoring the information loss in the primary analysis provides a natural reference point.

Rubin’s variance estimator was subsequently evaluated against the information anchored variance in the reference based and δ -adjusted MI settings. We now outline our methodological findings for each approach. We then compare and contrast our approach to variance estimation in the sensitivity analysis context with alternative proposals put forward for variance estimation following reference based analysis.

7.2 Reference based sensitivity analysis via multiple imputation

In Chapter 2 the reference based MI procedures of Carpenter, Roger and Kenward [13] were introduced as a novel approach to sensitivity analysis in the RCT setting. Missing values are imputed by reference to other patient groups in the trial. Consequently the specification of potentially numerous sensitivity analysis parameters is not required.

However, borrowing information between trial arms produces unwanted behaviour in the empirical long-run sampling variance of the reference based treatment effect. The strong assumption that deviators behave exactly like those in an opposing reference arm creates a covariance between the estimated means of these two groups. This reduces the long-run sampling variance of the reference

based treatment estimate, below that seen under MAR (typically the primary analysis assumption). Equally unappealingly, the sampling variance of the reference based treatment estimate becomes smaller than what would be obtained, were the deviation data observed in the given scenario. Further it becomes smaller and smaller as more data are missing. A simulation study in Section 2.3 confirmed the undesirable behaviour of the long-run sampling variance of the reference based MI estimator (equivalent to that of the ML estimator). Clearly this behaviour indicates this is not the variance we should be targeting.

The same simulation study revealed Rubin's MI treatment estimator is unbiased and Rubin's variance estimator behaves differently and appears to have more desirable behaviour. Rubin's variance estimator is always larger than the estimated variance we would obtain had the deviation data been observed, fulfilling our first principle. Rubin's variance estimator also increases as the proportion of missing data increases. These findings form part of a letter published in the Journal of Biopharmaceutical Statistics in response to comments published by Seaman et al. on Carpenter, Roger and Kenward's original paper proposing the reference based sensitivity MI framework [13, 57, 17].

In Chapter 3 we evaluated Rubin's variance estimator against our information anchoring principle. That is for sensitivity analysis which retains the primary design based analysis model and varies only the missing data assumption, the information loss due to missing data should be kept constant across primary and all sensitivity analysis. Exploration in a simple baseline and single follow-up setting, where Rubin's variance estimator could be derived exactly, revealed Rubin's variance estimator for the treatment effect, approximates the information anchored variance very closely. Simulation studies confirmed the approximation is excellent in a variety of realistic trial scenarios with baseline and single follow-up.

This led to a general theorem on the information anchoring properties of Rubin's variance estimator in the baseline and single follow-up trial setting in Chapter 4. The general result was then extended for the longitudinal trial setting, first with missingness at the final time point only and then with a more general monotone pattern of missingness in both arms. That is,

For all scenarios where the design based variance estimator, were we able to observe the post-deviation data under the specified de-facto scenario, can be expressed as the design based variance estimator were we able to observe all the post-deviation data under on-treatment/MAR behaviour plus a small component which is $O(n^{-2})$, Rubin's variance estimator is to an excellent approximation, information anchoring.

Simulations again confirmed that the approximation is excellent in a variety of realistic clinical trial settings. It is only with very high proportions of missing data or considerably different proportions of missing data across arms ($> 40\%$ in both arms or $> 50\%$ in one arm and 0% in the other) when the approximation is not so sharp. Depending on the trial scenario, results show there may be more or less information on the treatment estimator i.e. a greater or smaller variance estimate with high proportions of missing data, in comparison to the information anchored variance. The magnitude and direction of the difference depends on the difference between the variance of the imputation parameters that would be used under MAR and in the particular de facto setting of question.

Indeed with 50% or more missing data a greater amount of information is unobserved than observed, therefore one would be cautious of over interpreting the resulting treatment effect and variance. The analysis depends on untestable assumptions for more patients than are observed. Therefore the resulting treatment effect may not be so reliable in itself, as well as the variance. Therefore whilst the presented techniques can be used with more than 50% missing data, in the same way as standard MI under MAR [19, 44], in such circumstances we advise results should be presented with extreme caution. Just as we would with any other method of statistical analysis in the presence of such high proportions of data, which will also result in a strongly assumption-led treatment estimate and variance.

With large proportions of missing data ($> 50\%$) a detailed investigation into why there is so much missingness should be undertaken. The results of this investigation could indeed be more useful than the results of any analysis based on a strong untestable assumption for more than half the patients, which are likely to be biased. For example, perhaps an experimental drug was not well tolerated. Reasons why this occurred may help the drugs development more than a highly assumption dependent treatment estimate.

In practice however, such high proportions of missing data are unlikely in well designed trials. The results therefore provide a solid justification for using Rubin's treatment and variance estimate in reference based sensitivity analysis.

Chapter 6 illustrated just how useful and easily accessible the evaluated methodology is. Reference based sensitivity analysis for two RCT's was conducted using the newly developed Stata command, `mimix`. The analyses demonstrated the information anchoring properties of Rubin's variance estimator in real life settings.

7.3 The ' δ -method'

In Chapter 5 attention focussed on the ' δ -method' for sensitivity analysis in the RCT setting. Missing values are essentially imputed under MAR but then adjusted by δ to reflect a worse or better response post-deviation.

The ' δ -method' shares some similarities with the reference based approach; the mean of the MAR imputation distribution is altered to reflect departures from MAR. However information that is external to the main trial data set is used to alter the imputed values, via sensitivity analysis parameter δ , rather than borrowing information from opposing trial arms.

As we are not borrowing information from within the trial data, the empirical long-run sampling variance of the δ -adjusted MI estimate does not exhibit the strange behaviour observed in the reference based setting. The long-run sampling variance does not violate our first principle so markedly. Simulations confirmed this. However, as the primary design based analysis model is retained in sensitivity analysis we still desire the information anchored variance estimate in the δ -adjusted context. That is we desire to keep the information loss due to missing data constant across primary and sensitivity analysis.

Exploration in a simple baseline and single follow-up trial setting where Rubin's treatment effect

and variance estimator could be derived exactly revealed Rubin’s treatment effect is unbiased and the variance estimator does approximately preserve the information loss following δ -adjustment. Simulations confirmed the theoretical results. This led to a general theorem on the information anchoring properties of Rubin’s variance in the δ -adjusted baseline and single follow-up setting and more complex longitudinal setting. For fixed δ -adjusted sensitivity analysis, Rubin’s variance estimator is to an excellent approximation, information preserving. A better approximation is obtained than in the reference based case. Simulations confirmed the approximation is excellent in realistic trial settings, and more unrealistic settings with very high proportions of missing data.

When a prior distribution is placed on δ in imputation, i.e. for imputation k , $\delta_k \sim N(\delta, \sigma_\delta^2)$, the sharpness of the approximation depends on the variance of δ_k , σ_δ^2 . Simulations demonstrated that Rubin’s variance estimator provides a sharper information anchored variance for smaller σ_δ^2 . In Chapter 6 we conducted δ -adjusted sensitivity analysis for a peer review study. Results showed the excellent information anchoring property of Rubin’s variance estimator in fixed δ -adjusted sensitivity analysis. The results also demonstrated how the assumed variance for δ affects the information anchoring performance of Rubin’s variance estimator when this is not fixed at zero. The resulting variance estimate is not necessarily information anchoring. This illustrates just how important this parameter is when designing sensitivity analysis and how careful elicitation of the sensitivity parameters should be undertaken. To be information anchoring in δ -adjusted sensitivity analysis a variety of fixed δ adjustments should be made over plausible values of δ .

Rubin’s variance estimator, which also matches the empirical long-run sampling variance of the MI estimate in the δ -adjusted setting, is approximately information anchoring. It is encouraging that Rubin’s variance estimator, the information anchored variance and the long-run sampling variance of the MI estimate all agree in the δ -adjusted setting. In the reference based setting where the long-run sampling variance of the MI estimate is unnatural, due to the imputation manipulation, we are obtaining a variance estimate which has the same appropriate information anchoring properties.

We note that whilst the ‘ δ -method’ can also be used with more than 50% missing data, in the same way as standard MI under MAR, we advise results should be interpreted with extreme caution in such cases. As discussed in the previous section, results will be highly dependent on assumptions with $> 50\%$ missing data.

7.4 Generalisability

The theoretical results in Chapters 3–5 demonstrate unequivocally how, for the treatment effect of interest, Rubin’s MI treatment effect is unbiased and Rubin’s MI variance estimate provides an excellent approximation for the information anchored variance. The results apply to two arm baseline and single follow-up trial settings and longitudinal settings with either non-response at the final follow-up only or monotone missingness patterns, where interest lies in the treatment effect at the final follow-up. The general nature of the results are a key strength.

We focused on the longitudinal setting with monotone missingness since this is the most commonly observed pattern of non-response in longitudinal trials. In practice non-monotone missing data patterns may occur, for example where patients miss a planned follow-up visit but are later observed. When interest lies in the treatment effect at the final follow-up, as considered in this

thesis and is most often the case, the addition of interim missing data will not have a large impact, since interim missing data values will not affect the resulting treatment estimate. Further extensions of the results reported here are however required for longitudinal trials where interest lies in an alternative measure of the treatment effect, recorded over the course of the trial, e.g. an area under the curve and we are not prepared to assume MAR (we note if interim missing data are imputed under MAR the current results would not be impacted). If $V_{\text{DF, full}} = V_{\text{DJ, full}} + O(n^{-2})$ or indeed a component which is smaller than $O(n^{-2})$ it is plausible that the information anchoring argument we present can be extended to this situation.

Throughout this thesis we assumed the outcome was continuous and approximately normally distributed. It is also plausible that the information anchoring argument we present can be extended for other types of data. Again if $V_{\text{DF, full}} = V_{\text{DJ, full}} + O(n^{-2})$ holds and an appropriate transformation can be obtained to establish normality then results may naturally follow. For Generalized Linear Models (GLM), if we perform reference based imputation on the linear predictor scale, then we can apply the theory developed here on the linear predictor scale. This suggests that for GLMs, reference based simulation will be approximately information anchored. We do note, however, that issues may arise with non-collapsability when combining the component models in this setting which need investigation.

We also focused on the two arm trial setting. For trials with more than two arms it is natural how the results in Chapter 4 will extend for each treatment group comparison, again providing the key condition $V_{\text{DF, full}} = V_{\text{DJ, full}} + O(n^{-2})$ holds. For different trial designs additional components may be involved in the difference between Rubin’s variance estimate and the information anchoring variance. Careful exploration of the theory outlined in Chapter 4 would need to be undertaken in the specific scenario to establish the precise difference. For example, in a cluster randomised trial there is not only individual level variability but between cluster variability. It is conceivable that the information anchoring performance will depend on the difference in the between cluster variability by arm, as well as the individual level variability by arm with deviation in both arms.

Following the inference in this thesis we envisage the difference between Rubin’s variance and the information anchoring variance will not be largely impacted in alternative scenarios for realistic proportions of missing data i.e. $< 50\%$ missing. The within variance component of Rubin’s variance estimate, which averages the estimated treatment variance over the imputation set asymptotically approaches the variance of the treatment estimator were we able to observed the post deviation data. It is the proportion of missing data combined with the difference in the between imputation variance of the treatment estimator under MAR imputation and the required de facto imputation which impacts the information anchoring performance. For realistic amounts of missing data ($< 50\%$) the combined impact of these factors is not anticipated to be too large, relative to the information anchored variance we desire.

It was important to explore the derived results in realistic trial settings, inspired by real data. The two case studies analysed in Chapter 6 enabled the information anchoring properties of Rubin’s variance to be observed in real life scenarios. The asthma trial also provided a suitable structure for the simulation investigations. Using a real dataset to provide structure avoids arbitrary choices and aids the generalisability of the result. We note that the simulations were not an exact replica of the original study but a simplified version, with similarities for an underlying realistic framework.

Simulations inspired by the asthma trial confirmed the theoretical findings in a variety of real life scenarios. We also considered more extreme scenarios i.e. higher/lower covariance structures with

high amounts of missingness to demonstrate the generalisability of the results. While we were limited in the number of scenarios we could explore within simulations, we were able to assess the approximation in a number of the most likely settings to support the theoretical findings.

Throughout this thesis the controlled imputation methods we have investigated assume the primary analysis is conducted under MAR. This assumption is often most plausible and departures from MAR can be better evaluated via sensitivity analysis. It can be hard to justify a particular MNAR model. Analysis under MAR does not require additional modelling of the dropout procedure. Therefore MAR provides a natural starting point for subsequent MNAR exploration.

As explained by Carpenter, Roger, Kenward and myself in [57] “we envisage the typical situation where a patient attends a scheduled clinic visit at which —after undergoing the planned clinical measurements, discussing their progress with the clinical team and in consultation with the clinical team— the patient decides to deviate from the protocol in a particular manner. Depending on the clinical context the nature of these deviations will vary widely. Examples might include the addition of rescue medication, reduction or discontinuation of the study drug, etc. In some studies this deviation may be determined entirely by the clinical measurements recorded in the trial database. For example, in a blood pressure reduction trial the patient’s blood pressure may have exceeded a predefined level triggering a change of drugs. In such a setting the reason for deviation depends only on fully observed data, so we can talk of deviation at random (defined by analogy to missing at random, noting that this is conditional on that patient’s observed data).”

However, if this is not the case the reason for deviation may depend on unknown data and consequently post-deviation data will not be MAR. If there are strong reasons to suggest otherwise then one might consider that primary analysis under a pre-specified MNAR assumption is preferable. One of the MNAR modelling approaches discussed in Section 1.3 could be employed to do so. Any single MNAR analysis however depends heavily on an assumption which is not testable.

As no MNAR model can be considered definitive several MNAR models will still be required to assess sensitivity. In such cases MAR can still be a suitable approximation assumption for estimating parameters used later to impute post-deviation data for sensitivity analysis. The observed data alone does not tell us anything about the missing data, so MAR provides a natural starting point for subsequent alternative MNAR exploration. Therefore the use of the controlled imputation methods discussed in this thesis can still be appropriate if reference based or δ based assumptions are contextually relevant.

In all trials it is important that trialists only employ methods which make plausible assumptions for the setting at hand. We acknowledge that in certain circumstances the assumptions made by the reference based MI procedures may not be appropriate. For example, consider a surgical trial comparing standard ankle replacement versus ankle fusion for end stage ankle arthritis. Patients are randomised on the day of surgery to ensure minimal drop out pre-surgery. Planned follow-up visits occur at 3 and 6 months post-surgery. In this setting it may not be appropriate to assume that patients who deviate at 3 months post-surgery jump to follow the behaviour in the other arm. Unlike in a drug trial the treatment schedule has already been completed, therefore this would not be a realistic scenario. An assumption that the deviators experience an outcome that is a specified amount worse than those observed could however be more realistic.

We advise trialists to carefully consider the assumptions behind the controlled multiple imputation analyses to ensure each analysis undertaken is appropriate to the context at hand. Indeed the

assumptions behind any method of statistical analysis should be relevant to the context at hand.

7.5 Information anchoring

For both of the evaluated controlled imputation approaches we have shown Rubin's MI variance estimator preserves the loss of information due to missing data in the primary analysis in the sensitivity analysis to an excellent approximation. Rubin's variance formula fulfills our information anchoring principle which we believe is crucial in sensitivity analysis which retains the primary design based analysis model.

Rubin's variance estimator has generally been accepted for use in the δ -adjusted MI setting. We have been unable to find any specific literature criticising its use or proposing otherwise. Since Rubin's variance estimator preserves the loss of information in the primary analysis in the δ -adjusted setting, this further supports the justification for Rubin's variance estimator in the more recently proposed reference based settings, where Rubin's variance estimate also has the same information anchoring properties.

There cannot be more information on the treatment effect in the sensitivity analysis than when the assumption is actually true, i.e. the deviation data is actually observed under the given scenario. This principle is fundamental to the type of sensitivity analysis proposed. As discussed, the conventional long-run sampling variance of the reference based treatment estimate does not meet this principle. Thus it is inappropriate.

Alongside the production of this thesis, the differential behaviour between Rubin's MI variance estimate and the conventional long-run sampling variance of the reference based treatment effect has been noted by Lu (2014) [76], Ayele et al. (2014) [77] and Seaman, White and Leacy (2014) [17]. As discussed in Chapter 2 the borrowing of data from other trial arms results in a smaller long-run sampling variance than the information anchored variance from Rubin's variance estimator. Consequently, confidence interval coverage greater than the nominal coverage rate is obtained by Rubin's variance estimate. This has therefore prompted the development of alternative implementations of the reference based pattern mixture modelling approach, which obtain variance estimates that match the long-run sampling variance of the reference based treatment effect.

Lu [76] introduced an analytical approach for the placebo based i.e. CR pattern mixture model as an alternative to MI. He derived an analytical expression for the CR treatment effect based on the parameters of the underlying MAR model using the product of conditional normal regression models. A posterior simulation or a ML based approach to inference via the delta method for variance estimation was proposed. Simulations confirmed that the resulting variance estimates were consistent with the long-run sampling variance of the CR treatment estimate.

Recently in 2015 Tang [78] derived different but equivalent analytical expressions for the placebo based (CR) pattern mixture model treatment effect and its variance via the likelihood based approach. Rather than factoring the observed (MAR) data likelihood in its conditional form as Lu did, Tang's formula were derived in matrix form from the likelihood of the mixed effects model for repeated measures (MMRM).

Both Lu and Tang’s formula are quite complex and do require special programing for each specific analysis. Further these approaches are only available for the CR method.

Liu and Pang in 2015 [79] proposed a Bayesian method for reference based pattern mixture modelling. They show the reference based treatment estimates for CIR, J2R and CR can be expressed as linear combinations of the parameters derived from a MMRM analysis via a Bayesian approach. Therefore the reference based treatment effect and its variance can can be obtained from MCMC samples based on the MMRM analysis. Or an approximate pattern mixture model analysis can be applied for variance estimation. A simulation study showed the variance estimates obtained from the full Bayesian approach and approximate pattern mixture method coincided with the long-run sampling variance for the CIR, J2R and CR reference based MI estimators.

The implementations of Lu, Tang and Liu and Pang all provide variance estimators which coincide with the conventional long-run sampling variance of the reference based treatment effect (which coincides with the conventional long-run sampling variance of the reference based MI treatment effect). They do not consider the fact that they are obtaining a variance estimate which is actually lower than if the reference based deviation data could be observed. They all miss this crucial point. A sensitivity analysis cannot lead us to knowing more about a situation than we would do if that situation was observed. We therefore do not believe their implementations are appropriate within the context of sensitivity analysis, nor even in primary analysis, because they do not acknowledge they are borrowing information.

We believe careful consideration still needs to be given in different contexts as to whether a variance estimate that is less than if the deviation data were truly observed, under the given reference based scenario, is actually meaningful.

7.6 Future work

The general results in Chapter 4 reveal that as long as the variance estimator for the treatment effect in the design based analysis where deviations occur off-treatment but the deviation data is fully observed ($V_{DF, full}$) can be expressed as the variance estimator for the treatment effect in the design based analysis with fully observed data if no deviations occur ($V_{DJ, full}$) plus a $O(n^{-2})$ term, then Rubin’s rules will provide an appropriate variance estimate which approximately preserves the information loss in the primary analysis. There is thus great scope to extend the MI options proposed by Carpenter, Roger and Kenward [13] and develop further contextually relevant sensitivity analyses framed around key clinical assumptions of interest. A combination of a reference based and δ -adjusted approach may be desirable in the trial setting, whereby deviators are postulated to have a worse/better response than a reference group by an amount δ .

For settings where $V_{DF, full} = V_{DJ, full} + O(n^{-2})$ does not follow, but one can specify an alternative relationship between $V_{DF, full}$ and $V_{DJ, full}$ we believe the theory in Chapter 4 can be readily adapted to establish the information anchoring performance of Rubin’s variance in that setting. Once the form of the alternative relationship between $V_{DF, full}$ and $V_{DJ, full}$ is known the general theoretical framework can be followed with the alternative relationship in place of $V_{DF, full} = V_{DJ, full} + O(n^{-2})$ to see how strong the information anchoring is. The general theoretical set-up provides a framework for future exploration in other specific settings.

Within this thesis we have looked specifically at variance estimation following reference based MI in the context of sensitivity analysis addressing departures from a primary MAR assumption. We acknowledge that further investigation of the variance estimator is required outside the context of sensitivity analysis. Particularly if reference based MI is to be committed for a primary analysis. As detailed in Section 7.5, other researchers have introduced alternative approaches for reference based pattern mixture modelling which target the conventional long-run variance of the reference based treatment effect. But we have demonstrated this variance is lower than the variance obtained if the reference based assumption were true and we observed the post-deviation data. We are not sure when it would be appropriate to have more information in an analysis with partially observed data based upon a strong assumption than when the assumption is actually true and all data is observed. This needs to be carefully explored.

If reference based MI is employed for a primary analysis some penalisation for missingness may alternatively be desired, but one might not want to impose the same penalisation for missingness as observed under MAR. For example, a further inflation of the variance may be required to account for the uncertainty. A loss of information is inevitable with missing data however more research is required to establish the degree of penalisation required outside the context of sensitivity analysis addressing departures from MAR. It is most conceivable that this will be situation specific and based on prior knowledge.

We have seen that in using the observed reference data to build reference based imputation models Rubin's variance formula provides an MAR information preserving variance estimate for the standard reference based MI approach. In Chapter 4 we identified how the size of the reference sample used to build the reference based imputation models could effect the resulting variance of the treatment effect. Specifically in Subsection 4.1.4 we used this approach to obtain an improved information anchoring variance estimate in more extreme trial scenarios with deviation in one arm. The sample size for the imputation model influences the precision of the imputed data, hence the overall treatment estimate, thus indicates a simple route to control the desired information loss.

A re-scaling of the variance of the parameters in the reference based approach could achieve the same objective. For example, in some cases it may be conceivable that we wish to inflate the variance by an extra 5% to allow for the information loss. The information anchoring approach described here provides a natural starting point for discussions.

A more developed Bayesian modelling approach that allows moderation of the penalisation imposed for missing data could potentially be desirable to formulate such an analysis. A full Bayesian approach could formally influence how much weight is given to the observed reference data in constructing the reference based pattern mixture models for the deviators –used for imputation or directly in a full Bayesian analysis – to control the information loss as desired. We briefly outline some further thoughts on how this may be achieved.

The power prior has been introduced as a general class of priors that enables weighting of historical data [80, 81]. As detailed by Ibrahim [80], the power prior distribution is constructed by raising the likelihood function of the historical data to the power a_0 , where $0 \leq a_0 \leq 1$. The parameter a_0 can essentially be interpreted as a precision parameter for the historical data. It quantifies the uncertainty in the historical data by controlling the heaviness of the tails of the prior distribution. When performing reference based analysis the reference data is essentially used twice. In a full Bayesian modelling approach to reference based analysis the observed reference data is also used to form the prior for the missing observations. By alternatively employing a power prior, based on

the reference group data for the missing observations, raised to the power a_0 and altering a_0 one could weight the influence of the observed reference data in the model for the deviators. Thus in a principled manner, control the penalisation for missingness.

These very initial thoughts require much further exploration but suggest a potential for a more structured route for variance estimation outside the current context of sensitivity analysis. However, the literature [80, 81] acknowledges that the power, α_0 , cannot be estimated from the data. It requires an external assumption. The specification of α_0 is essentially the same as choosing how much larger or smaller the variance of the imputed data should be, via a re-scaling or bootstrapping approach which ultimately have the same goal.

7.7 Concluding remarks

In conclusion we find a strong theoretical justification for using Rubin's MI combining rules following reference and δ based MI for sensitivity analysis of clinical trials with missing data. Rubin's MI treatment estimator is unbiased. Rubin's variance estimator anchors the information loss due to missing data over the primary (MAR) and sensitivity analysis to an excellent approximation. We have introduced a Stata program which implements the reference based MI procedures (`mimix`) and performed sensitivity analysis of two RCTs to illustrate the relevance and accessibility of the approach.

The evaluated controlled imputation techniques address the need for understandable and applicable methods of sensitivity analysis in the RCT setting, which was highlighted in the recent regulatory guidelines from the EMA [11] and an FDA-mandated panel report from the US National Research Council [12].

We hope this thesis will help facilitate more relevant accessible sensitivity analyses for busy trialists across the medical field. Ultimately ensuring appropriate inferences can be drawn from clinical trials, for improved patient care, despite the inevitable occurrence of missing data.

Appendix A

Exploratory simulation study details

We present the values of the α 's used in the models for the response in the exploratory simulation study introduced in Section 2.3. For each missing data scenario there is a model for the response at time point j , for $j = 2, 3, 4$. The value of α_0 was calibrated to achieve the required missingness.

Table A.1: Values of α in the model for response in the exploratory simulation study

Proportion of missing data at week 12 (j=4)	j	α_0	α_1	α_2	α_3
10%	2	0.155	1.705	-0.213	0.866
	3	-0.212	1.362	-0.077	0.318
	4	-0.091	1.201	-0.366	0.150
20%	2	0.782	1.705	-0.213	0.866
	3	-1.067	1.362	-0.077	0.318
	4	-0.462	1.201	-0.366	0.150
30%	2	1.312	1.705	-0.213	0.866
	3	-1.790	1.362	-0.077	0.318
	4	-0.774	1.201	-0.366	0.150
40%	2	1.705	1.705	-0.213	0.866
	3	-2.327	1.362	-0.077	0.318
	4	-1.006	1.201	-0.366	0.150

Appendix B

Reference based computations; baseline and single follow-up

B.1 The design based variance estimator when post-deviation data is observed

Here we present additional details of the calculations involved in establishing the expectation of the design based variance where post-deviation data is observed in Section 2.2.2. In the given setting, as presented in the main text we require the expectation of,

$$\frac{\hat{\sigma}_{22,r}}{n_r} + \frac{\hat{\sigma}_{22,a}}{n_a} = \frac{\frac{1}{n_r-1} \sum_{i=1}^{n_r} (Y_{ri2} - \bar{Y}_{r2})^2}{n_r} + \frac{\frac{1}{n_a-1} \sum_{i=1}^{n_a} \left(Y_{ai2} - \frac{n_o}{n_a} \bar{Y}_{a2,o} - \frac{n_d}{n_a} \bar{Y}_{a2,d} \right)^2}{n_a}.$$

For the first term we note,

$$E \left[\frac{1}{n_r-1} \sum_{i=1}^{n_r} (Y_{ri2} - \bar{Y}_{r2})^2 \right] = \sigma_{r22}.$$

Expanding the square and taking expectation term by term for the numerator of the second term gives,

$$E \left[\sum_{i=1}^{n_a} Y_{ai2}^2 \right] = n_o (\sigma_{a22} + \mu_{a2}^2) + n_d (\sigma_{d22} + \mu_{d2}^2),$$

$$E \left[\sum_{i=1}^{n_a} \left(\frac{n_o}{n_a} \right)^2 \bar{Y}_{a2o}^2 \right] = n_a \left(\frac{n_o}{n_a} \right)^2 \left(\frac{\sigma_{a22}}{n_o} + \mu_{a2}^2 \right),$$

$$E \left[\sum_{i=1}^{n_a} \left(\frac{n_d}{n_a} \right)^2 \bar{Y}_{a2d}^2 \right] = n_a \left(\frac{n_d}{n_a} \right)^2 \left(\frac{\sigma_{d22}}{n_d} + \mu_{d2}^2 \right),$$

$$E \left[\sum_{i=1}^{n_a} -2 \frac{n_o}{n_a} Y_{ai2} \bar{Y}_{a2o} \right] = -2 \frac{n_o^2}{n_a} \left[\frac{\sigma_{a22}}{n_o} + \mu_{a2}^2 \right] - 2 \frac{n_o n_d}{n_a} \mu_{a2} \mu_{d2},$$

$$E \left[\sum_{i=1}^{n_a} -2 \frac{n_d}{n_a} Y_{ai2} \bar{Y}_{a2d} \right] = -2 \frac{n_o n_d}{n_a} \mu_{a2} \mu_{d2} - 2 \frac{n_d^2}{n_a} \left[\frac{\sigma_{d22}}{n_d} + \mu_{d2}^2 \right],$$

$$E \left[\sum_{i=1}^{n_a} 2 \frac{n_o n_d}{n_a^2} \bar{Y}_{a2o} \bar{Y}_{a2d} \right] = 2 \frac{n_o n_d}{n_a} \mu_{a2} \mu_{d2}.$$

Combining these terms gives,

$$E \left[\frac{1}{n_a - 1} \sum_{i=1}^{n_a} \left(Y_{ai2} - \frac{n_o}{n_a} \bar{Y}_{a2o} - \frac{n_d}{n_a} \bar{Y}_{a2d} \right)^2 \right] = \frac{1}{n_a - 1} \left[(n_a - 1) \left(\frac{n_o}{n_a} \sigma_{a22} + \frac{n_d}{n_a} \sigma_{d22} \right) + \frac{n_o n_d}{n_a} \Delta^2 \right],$$

where $\Delta = \mu_{a2} - \mu_{d2}$. Combining this result with the expectation for the first term gives (2.9) as required.

Assuming that n_a and n_r are sufficiently large so that we may take $(n_a - 1)$ to be n_a and $(n_r - 1)$ to be n_r and $n_a = n_r = n$ and $\sigma_{a22} = \sigma_{r22} = \sigma_{d22} = \sigma_{22}$ i.e. assuming equally sized trial arms with equal variance structure, the expected variance of the treatment estimate is,

$$\frac{2\sigma_{22}}{n} + \frac{n_o n_d}{n^3} \Delta^2.$$

Suppose that we also see deviation in the reference arm. That is only $n_{r,o}$ of the n reference arm patients ($i = r$) are observed at both times without deviation. Post-baseline the remaining $n_{r,d}$ reference arm patients deviate such that $n_{r,o} + n_{r,d} = n$.

With deviation in the reference arm $\bar{Y}_{r2} = \frac{n_{r,o}}{n} \bar{Y}_{r2,o} + \frac{n_{r,d}}{n} \bar{Y}_{r2,d}$. Therefore for the reference arm we require,

$$E \left[\frac{1}{n-1} \sum_{i=1}^n \left(Y_{ri2} - \frac{n_{r,o}}{n} \bar{Y}_{r2,o} - \frac{n_{r,d}}{n} \bar{Y}_{r2,d} \right)^2 \right].$$

Following the steps outlined above for $\sigma_{a22} = \sigma_{r22} = \sigma_{d22} = \sigma_{d22,r} = \sigma_{22}$ where $n_r - 1 = n_r$, we obtain,

$$E \left[\frac{1}{n-1} \sum_{i=1}^n \left(Y_{ri2} - \frac{n_{r,o}}{n} \bar{Y}_{r2,o} + \frac{n_{r,d}}{n} \bar{Y}_{r2,d} \right)^2 \right] = \sigma_{22} + \frac{n_{r,o} n_{r,d}}{n^2} \Delta_r^2,$$

where $\Delta_r = \mu_{r2} - \mu_{d2,r}$. Thus the overall expected variance for the treatment effect with deviation in both arms will be,

$$\frac{2\sigma_{22}}{n} + \frac{n_o n_d}{n^3} \Delta^2 + \frac{n_{r,o} n_{r,d}}{n^3} \Delta_r^2.$$

For a typical RCT powered at 80% with 5% statistical significance $\Delta^2 \approx 15.68 \sigma_{22} / n$ using (3.17). Therefore $V_{DF, \text{full}} = V_{DJ, \text{full}} + O(n^{-2})$.

B.2 The design based variance estimator when post-deviation data is unobserved; maximum likelihood analysis result

We derive the variance of the treatment effect when post-deviation data is unobserved from a full ML analysis following the approach of factored likelihoods presented in [22].

For the active arm we have complete data (Y_{ai1}, Y_{ai2}) for $i = 1, \dots, n_o$ patients and incomplete data, (Y_{ai1}) for the remaining n_d cases. Anderson [82] shows that the joint distribution of Y_{ai1} and Y_{ai2} can be factored into the marginal distribution of Y_{ai1} and the conditional distribution of Y_{ai2} given Y_{ai1} as follows,

$$f(Y_{ai1}, Y_{ai2} | \mu_{\mathbf{a}}, \Sigma) = f(Y_{ai1} | \mu_{a1}, \sigma_{11}) f(Y_{ai2} | \beta_{20.1} + \beta_{21.1} Y_{ai1}, \sigma_{2.1}),$$

where $\beta_{20.1} = \mu_{a2} - \beta_{21.1} \mu_{a1}$, $\beta_{21.1} = \frac{\sigma_{12}}{\sigma_{11}}$ and $\sigma_{2.1} = \sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}$. Which implies, $\mu_{a2} = \beta_{20.1} + \beta_{21.1} \mu_{a1}$, $\sigma_{12} = \beta_{21.1} \sigma_{11}$ and $\sigma_{22} = \sigma_{2.1} \beta_{21.1}^2 \sigma_{11}$.

The data density can therefore be written as,

$$\prod_{i=1}^n f(Y_{ai1} | \mu_{a1}, \sigma_{11}) \prod_{i=1}^{n_o} f(Y_{ai2} | \beta_{20.1} + \beta_{21.1} Y_{ai1}, \sigma_{2.1}).$$

Since the parameters of the two factors are distinct i.e. (μ_{a1}, σ_{11}) and $(\beta_{20.1}, \beta_{21.1}, \sigma_{2.1})$ are distinct, ML estimates can be found by independently maximising the likelihoods which correspond to these two factors. We achieve this by differentiating the corresponding log likelihoods. Since log is a strictly increasing function the values of the parameters which maximise the likelihood will also maximise the log likelihood.

Maximising the log likelihood for the first component gives, $\hat{\mu}_{a1} = \frac{1}{n} \sum_{i=1}^n Y_{ai1} = \bar{Y}_{a1}$ and $\hat{\sigma}_{11} = \frac{1}{n} \sum_{i=1}^n (Y_{ai1} - \hat{\mu}_{a1})^2$.

The second component is a classical linear regression model so making use of standard regression results we obtain, $\hat{\beta}_{21.1} = \frac{s_{12}}{s_{11}}$, $\hat{\beta}_{20.1} = \frac{1}{n_o} \sum_{i=1}^{n_o} Y_{ai2} - \hat{\beta}_{21.1} \frac{1}{n_o} \sum_{i=1}^{n_o} Y_{ai1} = \bar{Y}_{a2,o} - \hat{\beta}_{21.1} \bar{Y}_{a1,o}$ and $\hat{\sigma}_{2.1} = s_{2.1}$ where $s_{lk} = \frac{1}{n_o} \sum_{i=1}^{n_o} (Y_{ail} - \bar{Y}_{al,o})(Y_{aik} - \bar{Y}_{ak,o})$ for $l, k = 1, 2$ and $s_{2.1} = s_{22} - \frac{s_{12}^2}{s_{11}}$.

The variance of the parameters are found by calculating and inverting the information matrix (I), which is itself found by differentiating the log likelihoods twice and taking the negative.

Differentiation of the first log likelihood with respect to (μ_{a1}, σ_{11}) gives,

$$I(\hat{\mu}_{a1}, \hat{\sigma}_{11}) = \begin{bmatrix} \frac{n}{\hat{\sigma}_{11}} & 0 \\ 0 & \frac{n}{2\hat{\sigma}_{11}^2} \end{bmatrix}.$$

Differentiation of the second log likelihood with respect to $(\beta_{20.1}, \beta_{21.1}, \sigma_{2.1})$ gives,

$$I(\hat{\beta}_{20.1}, \hat{\beta}_{21.1}, \hat{\sigma}_{2.1}) = \begin{bmatrix} \frac{n_o}{\hat{\sigma}_{2.1}} & \frac{n_o \bar{Y}_{a1,o}}{\hat{\sigma}_{2.1}} & 0 \\ \frac{n_o \bar{Y}_{a1,o}}{\hat{\sigma}_{2.1}} & \frac{\sum_{i=1}^{n_o} \bar{Y}_{a1,i}^2}{\hat{\sigma}_{2.1}} & 0 \\ 0 & 0 & \frac{n_o}{2\hat{\sigma}_{2.1}^2} \end{bmatrix}.$$

Inverting these matrices and combining gives the large-sample covariance matrix for $(\mu_{a1}, \sigma_{11}, \beta_{20.1}, \beta_{21.1}, \sigma_{2.1})$ which we define as C,

$$C = \begin{bmatrix} \frac{\hat{\sigma}_{11}}{n} & 0 & 0 & 0 & 0 \\ 0 & \frac{2\hat{\sigma}_{11}^2}{n} & 0 & 0 & 0 \\ 0 & 0 & \frac{\hat{\sigma}_{2.1} \left(1 + \frac{\bar{Y}_{a1,o}^2}{s_{11}}\right)}{n_o} & -\frac{\bar{Y}_{a1,o} \hat{\sigma}_{2.1}^2}{n_o s_{11}} & 0 \\ 0 & 0 & -\frac{\bar{Y}_{a1,o} \hat{\sigma}_{2.1}^2}{n_o s_{11}} & \frac{\hat{\sigma}_{2.1}^2}{n_o s_{11}} & 0 \\ 0 & 0 & 0 & 0 & \frac{2\hat{\sigma}_{2.1}^2}{n_o} \end{bmatrix}.$$

Of main interest is $\hat{\mu}_{a2}$. We recall $\hat{\mu}_{a2}$ is a function of $(\hat{\mu}_{a1}, \hat{\beta}_{20.1}, \hat{\beta}_{21.1})$, for which we already know the ML estimates.

We make use of the following result presented in [22]: *Let $g(\theta)$ be a function of the parameter θ . If $\hat{\theta}$ is a maximum likelihood estimate of θ then $g(\hat{\theta})$ is a maximum likelihood estimate of $g(\theta)$.*

Thus the parametrisation means the ML estimate $\hat{\mu}_{a2} = \hat{\beta}_{20.1} + \hat{\beta}_{21.1} \hat{\mu}_{a1} = \bar{Y}_{a2,o} - \hat{\beta}_{21.1} \bar{Y}_{a1,o} + \hat{\beta}_{21.1} \bar{Y}_{a1} = \bar{Y}_{a2,o} + \hat{\beta}_{21.1} (\bar{Y}_{a1} - \bar{Y}_{a1,o})$.

The variance of $\hat{\mu}_{a2}$ is found using the following result also presented in [22]: *Let $g(\theta)$ be a monotone differentiable function of θ and let C be the asymptotic covariance matrix of $\theta - \hat{\theta}$ then the large-sample covariance matrix of $g(\theta) - g(\hat{\theta})$ is $D_g(\hat{\theta}) C D_g(\hat{\theta})^T$, where $D_g(\theta) = \frac{\partial g(\theta)}{\partial \theta}$ is the partial derivative of g with respect to θ .*

As $\mu_{a2} = \beta_{20.1} + \beta_{21.1} \mu_{a1}$, $D(\mu_{a2}) = (\hat{\beta}_{21.1}, 0, 1, \bar{Y}_{a1}, 0)$. C is given above.

Calculating $D(\mu_{a2}) C D(\mu_{a2})^T$ results in,

$$\hat{\sigma}_{2.1} \left[\frac{1}{n_o} + \frac{\hat{\rho}^2}{n(1-\hat{\rho}^2)} + \frac{(\bar{Y}_{a1,o} - \bar{Y}_{a1})^2}{n_o s_{11}} \right],$$

where $\hat{\rho}^2 = \frac{\sigma_{12}^2}{\sigma_{11} \sigma_{22}}$. Taking expectation and simplifying results in,

$$\frac{\sigma_{22}}{n_o} - \frac{\sigma_{12}^2}{\sigma_{11}} \frac{n_d}{n_o n} = \frac{\sigma_{22}}{n_o} - \frac{\sigma_{12}^2}{\sigma_{11}} \frac{\pi_d}{n(1-\pi_d)}.$$

For the reference arm $\mathbf{Y}_r = (Y_{ri1}, Y_{ri2})$ for $i = 1, \dots, n$. We can therefore use standard results

to establish $\hat{\mu}_{r2} = \frac{1}{n} \sum_{i=1}^n Y_{ri2} = \bar{Y}_{r2}$ with variance $\frac{\hat{\sigma}_{22}}{n}$. Taking expectation and combining this result with the variance for the active arm, with a simple addition since the two arms consist of independent patients, gives the result presented in the main text (3.7).

B.3 Reference based imputation models

To derive the reference based MI imputation models we utilise the properties of the conditional normal distribution which we state here for reference.

Conditional normal distribution

Consider a generic 2 x 1 vector $\mathbf{Y} = (Y_1, Y_2)^T$, which follows a bivariate normal distribution with mean $\mu = (\mu_1, \mu_2)^T$ and variance-covariance matrix,

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}.$$

Then the conditional distribution of $Y_2|Y_1$ is a multivariate normal distribution with mean,

$$\mu_2 + \frac{\sigma_{12}}{\sigma_{11}}(Y_1 - \mu_1) = \mu_2 + \beta_{21.1}(Y_1 - \mu_1) = \beta_{20.1} + \beta_{21.1}Y_1 \text{ and variance, } \sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}.$$

In the main text we have presented the conditional distributions for $\mathbf{Y}_{\mathbf{a}2, \mathbf{o}}|\mathbf{Y}_{\mathbf{a}1, \mathbf{o}}^*$ and $\mathbf{Y}_{\mathbf{r}2}|\mathbf{Y}_{\mathbf{r}1}^*$ and derived the ordinary least squares estimates of the included parameters, along with their variances. These ingredients, combined with the properties of the conditional normal distribution allow us to construct the required imputation models.

MAR

Under MAR with deviation following time 1 $\mathbf{Y} = (Y_{ai1}, Y_{ai2})^T$ for $i \in \mathcal{D}$ is postulated to follow a bivariate normal distribution with mean,

$$\mu = (\mu_{a1}, \mu_{a2})^T,$$

and variance,

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}.$$

So the conditional distribution of $Y_{ai2}|Y_{ai1}$ for $i \in \mathcal{D}$ is multivariate normal with mean,

$$\mu_{a2} + \beta_{21.1}(Y_{ai1} - \mu_{a1}),$$

and variance,

$$\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}.$$

For each imputation k we let the observed data dominate the prior and take a draw of the mean

parameters (the variance we assume is known) from the posterior, thus for imputation k , the conditional distribution of $Y_{ai2}|Y_{ai1}$ is multivariate normal with mean,

$$\hat{\mu}_{a2} + \hat{\beta}_{21.1} (Y_{ai1} - \hat{\mu}_{a1}).$$

Substituting in the ML estimate for $\hat{\mu}_{a2} = \hat{\beta}_{20.1} + \hat{\beta}_{21.1}\hat{\mu}_{a1}$ this is,

$$\begin{aligned} & \hat{\beta}_{20.1} + \hat{\beta}_{21.1}\hat{\mu}_{a1} + \hat{\beta}_{21.1} (Y_{ai1} - \hat{\mu}_{a1}) \\ &= \hat{\beta}_{20.1} + \hat{\beta}_{21.1}Y_{ai1}, \end{aligned}$$

which in substituting our posterior draws $\hat{\beta}_{20.1} = \bar{Y}_{a2,o} - \hat{\beta}_{21.1}\bar{Y}_{a1,o} + u_k$, and $\hat{\beta}_{21.1} = \frac{r}{q} + b_k$ and a flat improper prior for the missing observations themselves gives,

$$\begin{aligned} & \bar{Y}_{a2,o} - \left(\frac{r}{q} + b_k\right)\bar{Y}_{a1,o} + u_k + \left(\frac{r}{q} + b_k\right)Y_{ai1} + e_{i,k} \\ &= \bar{Y}_{a2,o} + u_k + \left(\frac{r}{q} + b_k\right)(Y_{ai1} - \bar{Y}_{a1,o}) + e_{i,k}, \end{aligned}$$

where,

$$\begin{aligned} u_k &\sim N(0, n_o^{-1}\sigma_{2.1}) \\ b_k &\sim N(0, q^{-1}\sigma_{2.1}) \\ e_{i,k} &\sim N(0, \sigma_{2.1}), \end{aligned}$$

for $r = \sum_{i \in \mathcal{O}} (Y_{ai1} - \bar{Y}_{a1,o})(Y_{ai2} - \bar{Y}_{a2,o})$ and $q = \sum_{i \in \mathcal{O}} (Y_{ri1} - \bar{Y}_{r1})^2$.

Copy reference

Under CR with deviation following time 1 $\mathbf{Y} = (Y_{ai1}, Y_{ai2})^T$ for $i \in \mathcal{D}$ is postulated to follow a bivariate normal distribution with mean,

$$\mu = (\mu_{r1}, \mu_{r2})^T,$$

and variance,

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}.$$

So the conditional distribution of $Y_{ai2}|Y_{ai1}$ for $i \in \mathcal{D}$ is multivariate normal with mean,

$$\mu_{r2} + \beta_{21.1} (Y_{ai1} - \mu_{r1}),$$

and variance,

$$\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}.$$

For each imputation k we let the observed data dominate the prior and take a draw of the mean parameters (the variance we assume is known) from the posterior, thus for imputation k , the conditional distribution of $Y_{ai2}|Y_{ai1}$ is multivariate normal with mean,

$$\hat{\mu}_{r2} + \hat{\beta}_{21.1} (Y_{ai1} - \hat{\mu}_{r1}).$$

Substituting in the ML estimate for $\hat{\mu}_{r2} = \hat{\beta}_{20.1} + \hat{\beta}_{21.1}\hat{\mu}_{r1}$ this is,

$$\begin{aligned} & \hat{\beta}_{20.1} + \hat{\beta}_{21.1}\hat{\mu}_{r1} + \hat{\beta}_{21.1}(Y_{ai1} - \hat{\mu}_{r1}) \\ &= \hat{\beta}_{20.1} + \hat{\beta}_{21.1}Y_{ai1}, \end{aligned}$$

which in substituting our posterior draws $\hat{\beta}_{20.1} = \bar{Y}_{r2} - \hat{\beta}_{21.1}\bar{Y}_{r1} + u_k$, and $\hat{\beta}_{21.1} = \frac{r}{q} + b_k$ and a flat improper prior for the missing observations themselves gives,

$$\begin{aligned} & \bar{Y}_{r2} - \left(\frac{r}{q} + b_k\right)\bar{Y}_{r1} + u_k + \left(\frac{r}{q} + b_k\right)Y_{ai1} + e_{i,k} \\ &= \bar{Y}_{r2} + u_k + \left(\frac{r}{q} + b_k\right)(Y_{ai1} - \bar{Y}_{r1}) + e_{i,k}, \end{aligned}$$

where,

$$\begin{aligned} u_k &\sim N(0, n^{-1}\sigma_{2.1}) \\ b_k &\sim N(0, q^{-1}\sigma_{2.1}) \\ e_{i,k} &\sim N(0, \sigma_{2.1}), \end{aligned}$$

for $r = \sum_{i=1}^n (Y_{ri1} - \bar{Y}_{r1})(Y_{ri2} - \bar{Y}_{r2})$ and $q = \sum_{i=1}^n (Y_{ri1} - \bar{Y}_{r1})^2$.

Jump to reference

Under J2R with deviation following time 1 $\mathbf{Y} = (Y_{ai1}, Y_{ai2})^T$ for $i \in \mathcal{D}$ is postulated to follow a bivariate normal distribution with mean,

$$\boldsymbol{\mu} = (\mu_{a1}, \mu_{r2})^T,$$

and variance,

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}.$$

So the conditional distribution of $Y_{ai2}|Y_{ai1}$ for $i \in \mathcal{D}$ is thus multivariate normal with mean,

$$\mu_{r2} + \beta_{21.1}(Y_{ai1} - \mu_{a1}),$$

and variance,

$$\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}.$$

For each imputation k we let the observed data dominate the prior and take a draw of the mean parameters (the variance we assume is known) from the posterior, thus for imputation k , the conditional distribution of $Y_{ai2}|Y_{ai1}$ is multivariate normal with mean,

$$\hat{\mu}_{r2} + \hat{\beta}_{21.1}(Y_{ai1} - \hat{\mu}_{a1}).$$

Substituting in the ML estimate for $\hat{\mu}_{r2}$ this is,

$$\hat{\beta}_{20.1} + \hat{\beta}_{21.1}\hat{\mu}_{r1} + \hat{\beta}_{21.1}(Y_{ai1} - \hat{\mu}_{a1}),$$

which after substituting in our posterior draws $\hat{\beta}_{20.1} = \bar{Y}_{r2} - \hat{\beta}_{21.1}\bar{Y}_{r1} + u_k$, $\hat{\beta}_{21.1} = \frac{r}{q} + b_k$, $\hat{\mu}_{r1} = \bar{Y}_{r1} + r_k$ and $\hat{\mu}_{a1} = \bar{Y}_{a1} + a_k$ and flat improper prior for the missing observations themselves gives,

$$\bar{Y}_{r2} - \left(\frac{r}{q} + b_k\right) \bar{Y}_{r1} + u_k + \left(\frac{r}{q} + b_k\right) (\bar{Y}_{r1} + r_k) + \left(\frac{r}{q} + b_k\right) Y_{ai1} - \left(\frac{r}{q} + b_k\right) (\bar{Y}_{a1} + a_k) + e_{i,k}.$$

Simplified this is,

$$\bar{Y}_{r2} + u_k + \left(\frac{r}{q} + b_k\right) (Y_{ai1} - \bar{Y}_{a1}) + \left(\frac{r}{q} + b_k\right) (r_k - a_k) + e_{i,k},$$

where,

$$\begin{aligned} u_k &\sim N(0, n^{-1}\sigma_{2.1}) \\ b_k &\sim N(0, q^{-1}\sigma_{2.1}) \\ a_k &\sim N(0, n^{-1}\sigma_{11}) \\ r_k &\sim N(0, n^{-1}\sigma_{11}) \\ e_{i,k} &\sim N(0, \sigma_{2.1}), \end{aligned}$$

for $r = \sum_{i=1}^n (Y_{ri1} - \bar{Y}_{r1})(Y_{ri2} - \bar{Y}_{r2})$ and $q = \sum_{i=1}^n (Y_{ri1} - \bar{Y}_{r1})^2$.

Copy increments in reference

Under CIR with deviation following time 1 $\mathbf{Y} = (Y_{ai1}, Y_{ai2})^T$ for $i \in \mathcal{D}$ is postulated to follow a bivariate normal distribution with mean and variance,

$$\begin{aligned} \boldsymbol{\mu} &= (\mu_{a1}, \mu_{a1} + \mu_{r2} - \mu_{r1})^T, \\ \boldsymbol{\Sigma} &= \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}. \end{aligned}$$

So the conditional distribution of $Y_{ai2}|Y_{ai1}$ for $i \in \mathcal{D}$ is thus multivariate normal with mean,

$$(\mu_{a1} + \mu_{r2} - \mu_{r1}) + \beta_{21.1} (Y_{ai1} - \mu_{a1}),$$

and variance,

$$\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}.$$

For each imputation k we again let the observed data dominate the prior and take a draw of the mean parameters (the variance we assume is known) from the posterior, thus for imputation k , the conditional distribution of $Y_{ai2}|Y_{ai1}$ is multivariate normal with mean,

$$(\hat{\mu}_{a1} + \hat{\mu}_{r2} - \hat{\mu}_{r1}) + \hat{\beta}_{21.1} (Y_{ai1} - \hat{\mu}_{a1}).$$

Substituting in the ML estimate for $\hat{\mu}_{r2} = \hat{\beta}_{20.1} + \hat{\beta}_{21.1}\hat{\mu}_{r1}$ this is,

$$\hat{\mu}_{a1} + \hat{\beta}_{20.1} + \hat{\beta}_{21.1}\hat{\mu}_{r1} - \hat{\mu}_{r1} + \hat{\beta}_{21.1} (Y_{ai1} - \hat{\mu}_{a1}).$$

Substituting in our posterior draws $\hat{\beta}_{20.1} = \bar{Y}_{r2} - \hat{\beta}_{21.1}\bar{Y}_{r1} + u_k$, $\hat{\beta}_{21.1} = \frac{r}{q} + b_k$, $\hat{\mu}_{r1} = \bar{Y}_{r1} + r_k$ and

$\hat{\mu}_{a1} = \bar{Y}_{a1} + a_k$ and flat improper prior for the missing observations themselves gives,

$$\begin{aligned} \bar{Y}_{a1} + a_k + \bar{Y}_{r2} - \left(\frac{r}{q} + b_k\right) (\bar{Y}_{r1}) + u_k + \left(\frac{r}{q} + b_k\right) (\bar{Y}_{r1} + r_k) - (\bar{Y}_{r1} + r_k) \\ + \left(\frac{r}{q} + b_k\right) (Y_{ai1} - \bar{Y}_{a1} - a_k) + e_{i,k}, \end{aligned}$$

which is,

$$\bar{Y}_{a1} + \bar{Y}_{r2} - \bar{Y}_{r1} + u_k + \left(\frac{r}{q} + b_k\right) (Y_{ai1} - \bar{Y}_{a1}) + \left(\frac{r}{q} + b_k\right) (r_k - a_k) + a_k - r_k + e_{i,k},$$

where,

$$\begin{aligned} u_k &\sim N(0, n^{-1}\sigma_{2.1}) \\ b_k &\sim N(0, q^{-1}\sigma_{2.1}) \\ a_k &\sim N(0, n^{-1}\sigma_{11}) \\ r_k &\sim N(0, n^{-1}\sigma_{11}) \\ e_{i,k} &\sim N(0, \sigma_{2.1}), \end{aligned}$$

for $r = \sum_{i=1}^n (Y_{ri1} - \bar{Y}_{r1}) (Y_{ri2} - \bar{Y}_{r2})$ and $q = \sum_{i=1}^n (Y_{ri1} - \bar{Y}_{r1})^2$.

Last mean carried forward

Under LMCF with deviation following time 1 $\mathbf{Y} = (Y_{ai1}, Y_{ai2})^T$ for $i \in \mathcal{D}$ is postulated to follow a bivariate normal distribution with mean and variance,

$$\begin{aligned} \mu &= (\mu_{a1}, \mu_{a1})^T, \\ \Sigma &= \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}. \end{aligned}$$

So the conditional distribution of $Y_{ai2}|Y_{ai1}$ for $i \in \mathcal{D}$ is thus multivariate normal with mean,

$$\mu_{a1} + \beta_{21.1} (Y_{ai1} - \mu_{a1}),$$

and variance,

$$\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}.$$

For each imputation k we let the observed data dominate the prior and take a draw of the mean parameters (the variance we assume is known) from the posterior, thus for imputation k , the conditional distribution of $Y_{ai2}|Y_{ai1}$ is multivariate normal with mean,

$$\hat{\mu}_{a1} + \hat{\beta}_{21.1} (Y_{ai1} - \hat{\mu}_{a1}).$$

Substituting in our posterior draws for $\hat{\mu}_{a1}$ and $\hat{\beta}_{21.1}$ and flat improper prior for the missing observations themselves this is,

$$\bar{Y}_{a1} + u_k + \left(\frac{r}{q} + b_k\right) (Y_{ai1} - \bar{Y}_{a1} - u_k) + e_{i,k},$$

where,

$$\begin{aligned} u_k &\sim N(0, n^{-1}\sigma_{11}) \\ b_k &\sim N(0, q^{-1}\sigma_{2.1}) \\ e_{i,k} &\sim N(0, \sigma_{2.1}), \end{aligned}$$

for $r = \sum_{i \in \mathcal{O}} (Y_{ai1} - \bar{Y}_{a1,o}) (Y_{ai2} - \bar{Y}_{a2,o})$ and $q = \sum_{i \in \mathcal{O}} (Y_{ai1} - \bar{Y}_{a1,o})^2$.

B.4 Imputation calculations

B.4.1 MAR imputation calculations

Here we present additional details of the calculations involved in establishing the expectation of Rubin's variance estimator in the MAR case in Section 3.1.1. As presented in the main text we require $E(V_{\text{DJ, MI}}) = E(\hat{W}) + (1 + \frac{1}{K}) E(\hat{B})$. First we consider $E(\hat{W})$,

$$E(\hat{W}) = E(\hat{\sigma}_k^2) = \frac{E(\hat{\sigma}_{22,r})}{n} + \frac{E(\hat{\sigma}_{22,a,k})}{n},$$

where,

$$\begin{aligned} (n-1) E(\hat{\sigma}_{22,r}) &= E\left(\sum_{i=1}^n (Y_{ri2} - \hat{\mu}_{r2})^2\right) = (n-1)\sigma_{22}, \\ (n-1) E(\hat{\sigma}_{22,a,k}) &= E\left[\sum_{i \in \mathcal{O}} (Y_{ai2} - \hat{\mu}_{a2,k})^2\right] + E\left[\sum_{i \in \mathcal{D}} (\hat{Y}_{ai2,k} - \hat{\mu}_{a2,k})^2\right]. \end{aligned}$$

Decomposing the two-right hand components separately, over the observed cases we have,

$$\begin{aligned} &E\left[\sum_{i \in \mathcal{O}} (Y_{ai2} - \hat{\mu}_{a2,k})^2\right] \\ &= E\left[\sum_{i \in \mathcal{O}} \left(Y_{ai2} - \frac{n_o}{n} \bar{Y}_{a2o} - \frac{n_d}{n} \bar{Y}_{a2d,k}\right)^2\right] \\ &= E\left[\sum_{i \in \mathcal{O}} \left(Y_{ai2} - \frac{n_o}{n} \bar{Y}_{a2o} - \frac{n_d}{n} \left[\bar{Y}_{a2o} + u_k + \left(\frac{r}{q} + b_k\right) (\bar{Y}_{a1d} - \bar{Y}_{a1o}) + \bar{e}_k\right]\right)^2\right] \\ &= E\left[\sum_{i \in \mathcal{O}} \left((Y_{ai2} - \bar{Y}_{a2o}) - \frac{n_d}{n} u_k - \frac{n_d}{n} \left(\frac{r}{q} + b_k\right) (\bar{Y}_{a1d} - \bar{Y}_{a1o}) - \frac{n_d}{n} \bar{e}_k\right)^2\right]. \end{aligned}$$

Expanding the square and taking expectation term by term results in the following non-zero components,

$$\begin{aligned} E\left[\sum_{i \in \mathcal{O}} (Y_{ai2} - \bar{Y}_{a2o})^2\right] &= (n_o - 1)\sigma_{22}, \\ E\left[\sum_{i \in \mathcal{O}} \left(\frac{n_d}{n}\right)^2 u_k^2\right] &= \left(\frac{n_d}{n}\right)^2 \sigma_{2.1}, \\ E\left[\sum_{i \in \mathcal{O}} \left(\frac{n_d}{n}\right)^2 \left(\frac{r}{q} + b_k\right)^2 (\bar{Y}_{a1d} - \bar{Y}_{a1o})^2\right] &= \left(\frac{n_d}{n}\right)^2 n_o \left[\frac{\sigma_{12}^2}{\sigma_{11}} + \frac{2\sigma_{2.1}}{(n_o - 1)}\right] \left[\frac{1}{n_d} + \frac{1}{n_o}\right], \\ E\left[\sum_{i \in \mathcal{O}} \left(\frac{n_d}{n}\right)^2 \bar{e}_k^2\right] &= \frac{n_d n_o}{n^2} \sigma_{2.1}. \end{aligned}$$

Over the deviators we have,

$$\begin{aligned}
& E \left[\sum_{i \in \mathcal{D}} \left(\hat{Y}_{ai2,k} - \hat{\mu}_{a2,k} \right)^2 \right] \\
&= E \left[\sum_{i \in \mathcal{D}} \left(\bar{Y}_{a2o} + u_k + \left(\frac{r}{q} + b_k \right) (Y_{ai1} - \bar{Y}_{a1o}) + e_{i,k} - \frac{n_o}{n} \bar{Y}_{a2o} - \frac{n_d}{n} \bar{Y}_{a2d,k} \right)^2 \right] \\
&= E \left[\sum_{i \in \mathcal{D}} \left(\frac{n_o}{n} u_k + \left(\frac{r}{q} + b_k \right) (Y_{ai1} - \bar{Y}_{a1o}) - \frac{n_d}{n} \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{a1o}) - \frac{n_d}{n} \bar{e}_k + e_{i,k} \right)^2 \right].
\end{aligned}$$

Expanding the square and taking expectation term by term results in the following non-zero components,

$$E \left[\sum_{i \in \mathcal{D}} \left(\frac{n_o}{n} \right)^2 u_k^2 \right] = \frac{n_o n_d}{n^2} \sigma_{2,1},$$

$$E \left[\sum_{i \in \mathcal{D}} \left(\frac{r}{q} + b_k \right)^2 (Y_{ai1} - \bar{Y}_{a1o})^2 \right] = \left[\frac{\sigma_{12}^2}{\sigma_{11}} + \frac{2\sigma_{2,1}}{(n_o - 1)} \right] n_d \left[1 + \frac{1}{n_o} \right],$$

$$E \left[\sum_{i \in \mathcal{D}} \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right)^2 (\bar{Y}_{a1d} - \bar{Y}_{a1o})^2 \right] = \left(\frac{n_d}{n} \right)^2 n_d \left[\frac{\sigma_{12}^2}{\sigma_{11}} + \frac{2\sigma_{2,1}}{(n_o - 1)} \right] \left[\frac{1}{n_d} + \frac{1}{n_o} \right],$$

$$E \left[\sum_{i \in \mathcal{D}} \left(\frac{n_d}{n} \right)^2 \bar{e}_k^2 \right] = \left(\frac{n_d}{n} \right)^2 \sigma_{2,1},$$

$$E \left[\sum_{i \in \mathcal{D}} e_{i,k}^2 \right] = n_d \sigma_{2,1},$$

$$E \left[\sum_{i \in \mathcal{D}} -2 \frac{n_d}{n} \left(\frac{r}{q} + b_k \right)^2 (\bar{Y}_{a1d} - \bar{Y}_{a1o}) (Y_{ai1} - \bar{Y}_{a1o}) \right] = -2 \frac{n_d^2}{n} \left[\frac{\sigma_{12}^2}{\sigma_{11}} + \frac{2\sigma_{2,1}}{(n_o - 1)} \right] \left[\frac{1}{n_d} + \frac{1}{n_o} \right],$$

$$E \left[\sum_{i \in \mathcal{D}} -2 \frac{n_d}{n} \bar{e}_k e_{i,k} \right] = -2 \frac{n_d}{n} \sigma_{2,1}.$$

Combining the results over the observed and deviating cases for time 1 a baseline where $\mu_{a1} = \mu_{r1}$ results in (3.4). For $(1 + \frac{1}{K}) E(\hat{B})$ we require,

$$E[\hat{B}] = \frac{1}{K-1} E \left[\sum_{k=1}^K \left(\hat{\theta}_{\text{DJ}, k} - \hat{\theta}_{\text{DJ}, \text{MI}} \right)^2 \right],$$

where,

$$\hat{\theta}_{\text{DJ}, k} = \frac{n_o}{n} \bar{Y}_{a2o} + \frac{n_d}{n} \left[\bar{Y}_{a2o} + u_k + \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{a1o}) + \bar{e}_k \right] - \bar{Y}_{r2},$$

$$\hat{\theta}_{\text{DJ}, \text{MI}} = \frac{1}{K} \sum_{k=1}^K \hat{\theta}_{\text{DJ}, k}.$$

That is,

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} u_k - \frac{n_d}{n} \bar{u} + \frac{n_d}{n} \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{a1o}) - \frac{n_d}{n} \left(\frac{r}{q} + \bar{b} \right) (\bar{Y}_{a1d} - \bar{Y}_{a1o}) + \frac{n_d}{n} \bar{e}_k - \frac{n_d}{n} \bar{e} \right)^2 \right].$$

Expanding the square and taking expectation by term gives the following non zero components,

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 u_k^2 \right] = K \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{2,1}}{n_o},$$

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right)^2 (\bar{Y}_{a1d} - \bar{Y}_{a1o})^2 \right] = K \left(\frac{n_d}{n} \right)^2 \left[\frac{\sigma_{12}^2}{\sigma_{11}} + \frac{2\sigma_{2,1}}{(n_o-1)} \right] \left[\frac{1}{n_d} + \frac{1}{n_o} \right],$$

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \bar{e}_k^2 \right] = K \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{2,1}}{n_d},$$

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \bar{u}^2 \right] = \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{2,1}}{n_o},$$

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + \bar{b} \right)^2 (\bar{Y}_{a1d} - \bar{Y}_{a1o})^2 \right] = K \left(\frac{n_d}{n} \right)^2 \left[\frac{\sigma_{12}^2}{\sigma_{11}} + \frac{\sigma_{2,1}}{(n_o-1)} \left(\frac{K+1}{K} \right) \right] \left[\frac{1}{n_d} + \frac{1}{n_o} \right],$$

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \bar{e}^2 \right] = \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{2,1}}{n_d},$$

$$E \left[\sum_{k=1}^K -2 \left(\frac{n_d}{n} \right)^2 u_k \bar{u} \right] = -2 \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{2,1}}{n_o},$$

$$\begin{aligned} E \left[\sum_{k=1}^K -2 \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right) \left(\frac{r}{q} + \bar{b} \right) (\bar{Y}_{a1d} - \bar{Y}_{a1o})^2 \right] \\ = -2K \left(\frac{n_d}{n} \right)^2 \left[\frac{\sigma_{12}^2}{\sigma_{11}} + \frac{\sigma_{2,1}}{(n_o-1)} \left(\frac{K+1}{K} \right) \right] \left[\frac{1}{n_d} + \frac{1}{n_o} \right], \end{aligned}$$

$$E \left[\sum_{k=1}^K -2 \left(\frac{n_d}{n} \right)^2 \bar{e}_k \bar{e} \right] = -2 \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{2,1}}{n_d}.$$

Combining the results gives (3.5).

Tables B.1 and B.2 summarise Rubin's derived variance versus simulated Rubin's variance for the MAR treatment effect for the asthma RCT inspired setting with,

$$\begin{pmatrix} Y_{ai1} \\ Y_{ai2} \end{pmatrix} \sim N \left\{ \begin{pmatrix} 2.0 \\ 2.2 \end{pmatrix}; \begin{bmatrix} 0.4 & 0.2 \\ 0.2 & 0.6 \end{bmatrix} \right\}, \quad i \in \mathcal{O},$$

$$\begin{pmatrix} Y_{ri1} \\ Y_{ri2} \end{pmatrix} \sim N \left\{ \begin{pmatrix} 2.0 \\ 1.9 \end{pmatrix}; \begin{bmatrix} 0.4 & 0.2 \\ 0.2 & 0.6 \end{bmatrix} \right\}, \quad i = 1, \dots, n.$$

These parameters, inspired by the asthma RCT, are also used to assess the derived results for Rubin's variance estimate in the other reference based and δ adjusted settings, see Tables B.3 to F.2.

Proportion of Missing data	Derived Variance	Average Variance	MCSE
10% (N=25)	0.00503	0.00503	0.000011
20% (N=50)	0.00532	0.00532	0.000012
30% (N=75)	0.00569	0.00568	0.000015
40% (N=100)	0.00618	0.00619	0.000019
50% (N=125)	0.00687	0.00690	0.000026
60% (N=150)	0.00792	0.00792	0.000037
70% (N=175)	0.00966	0.00976	0.000058

Table B.1: Derived Rubin’s variance versus simulated Rubin’s variance (averaged over 1000 simulations) for the MAR treatment effect in the asthma RCT inspired setting with $K=50$.

Proportion of Missing data	Derived Variance	Average Variance	MCSE
10% (N=25)	0.00423	0.00422	0.000009
20% (N=50)	0.00452	0.00454	0.000010
30% (N=75)	0.00489	0.00489	0.000014
40% (N=100)	0.00538	0.00540	0.000018
50% (N=125)	0.00607	0.00612	0.000025
60% (N=150)	0.00712	0.00721	0.000036
70% (N=175)	0.00886	0.00897	0.000057

Table B.2: Derived Rubin’s variance versus simulated Rubin’s variance (averaged over 1000 simulations) for the MAR baseline adjusted treatment effect in the asthma RCT inspired setting with $K=50$.

B.4.2 CR imputation calculations

Here we present additional details of the calculations involved in establishing the expectation of Rubin’s variance estimator in the CR case in Section 3.2.1. As presented in the main text we require $E(V_{\text{df, MI}}) = E(\hat{W}) + (1 + \frac{1}{K}) E(\hat{B})$. First we consider $E(\hat{W})$,

$$E(\hat{W}) = E(\hat{\sigma}_k^2) = \frac{E(\hat{\sigma}_{22,r})}{n} + \frac{E(\hat{\sigma}_{22,a,k})}{n}.$$

By breaking down the numerator we see,

$$(n-1) E(\hat{\sigma}_{22,r}) = E\left(\sum_{i=1}^n (Y_{ri2} - \hat{\mu}_{r2})^2\right) = (n-1)\sigma_{22},$$

$$(n-1) E(\hat{\sigma}_{22,a,k}) = E\left[\sum_{i \in \mathcal{O}} (Y_{ai2} - \hat{\mu}_{a2,k})^2\right] + E\left[\sum_{i \in \mathcal{D}} (\hat{Y}_{ai2,k} - \hat{\mu}_{a2,k})^2\right].$$

Decomposing the two-right hand components for the active arm separately, over the observed cases we have,

$$\begin{aligned}
& E \left[\sum_{i \in \mathcal{O}} (Y_{ai2} - \hat{\mu}_{a2,k})^2 \right] \\
&= E \left[\sum_{i \in \mathcal{O}} \left(Y_{ai2} - \frac{n_o}{n} \bar{Y}_{a2o} - \frac{n_d}{n} \bar{Y}_{a2d,k} \right)^2 \right] \\
&= E \left[\sum_{i \in \mathcal{O}} \left(Y_{ai2} - \frac{n_o}{n} \bar{Y}_{a2o} - \frac{n_d}{n} \left[\bar{Y}_{r2} + u_k + \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{r1}) + \bar{e}_k \right] \right)^2 \right] \\
&= E \left[\sum_{i \in \mathcal{O}} \left((Y_{ai2} - \bar{Y}_{a2o}) + \frac{n_d}{n} (\bar{Y}_{a2o} - \bar{Y}_{r2}) - \frac{n_d}{n} u_k - \frac{n_d}{n} \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{r1}) - \frac{n_d}{n} \bar{e}_k \right)^2 \right].
\end{aligned}$$

Expanding the square results in the following non-zero required expectations,

$$E \left[\sum_{i \in \mathcal{O}} (Y_{ai2} - \bar{Y}_{a2o})^2 \right] = (n_o - 1) \sigma_{22},$$

$$E \left[\sum_{i \in \mathcal{O}} \left(\frac{n_d}{n} \right)^2 (\bar{Y}_{a2o} - \bar{Y}_{r2})^2 \right] = \left(\frac{n_d}{n} \right)^2 \left[\sigma_{22} \left(1 + \frac{n_o}{n} \right) + n_o (\mu_{a2} - \mu_{r2})^2 \right],$$

$$E \left[\sum_{i \in \mathcal{O}} \left(\frac{n_d}{n} \right)^2 u_k^2 \right] = n_o \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{2,1}}{n},$$

$$\begin{aligned}
& E \left[\sum_{i \in \mathcal{O}} \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right)^2 (\bar{Y}_{a1d} - \bar{Y}_{r1})^2 \right] \\
&= \left(\frac{n_d}{n} \right)^2 n_o \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{2\sigma_{2,1}}{(n-1)\sigma_{11}} \right] \left[\sigma_{11} \left(\frac{1}{n_d} + \frac{1}{n} \right) + (\mu_{a1} - \mu_{r1})^2 \right],
\end{aligned}$$

$$E \left[\sum_{i \in \mathcal{O}} \left(\frac{n_d}{n} \right)^2 \bar{e}_k^2 \right] = \frac{n_d n_o}{n^2} \sigma_{2,1},$$

$$\begin{aligned}
& E \left[\sum_{i \in \mathcal{O}} -2 \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a2o} - \bar{Y}_{r2}) (\bar{Y}_{a1d} - \bar{Y}_{r1}) \right] \\
&= -2 \left(\frac{n_d}{n} \right)^2 n_o \frac{\sigma_{12}}{\sigma_{11}} \left(\frac{\sigma_{12}}{n} + (\mu_{a1} - \mu_{r1})(\mu_{a2} - \mu_{r2}) \right).
\end{aligned}$$

Over the deviators we have,

$$\begin{aligned}
& E \left[\sum_{i \in \mathcal{D}} \left(\hat{Y}_{ai2,k} - \hat{\mu}_{a2,k} \right)^2 \right] \\
&= E \left[\sum_{i \in \mathcal{D}} \left(\bar{Y}_{r2} + u_k + \left(\frac{r}{q} + b_k \right) (Y_{ai1} - \bar{Y}_{r1}) + e_{i,k} - \frac{n_o}{n} \bar{Y}_{a2o} - \frac{n_d}{n} \bar{Y}_{a2d,k} \right)^2 \right] \\
&= E \left[\sum_{i \in \mathcal{D}} \left(\frac{n_o}{n} (\bar{Y}_{r2} - \bar{Y}_{a2o}) + \frac{n_o}{n} u_k + \left(\frac{r}{q} + b_k \right) (Y_{ai1} - \bar{Y}_{r1}) \right. \right. \\
&\quad \left. \left. - \frac{n_d}{n} \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{r1}) - \frac{n_d}{n} \bar{e}_k + e_{i,k} \right)^2 \right].
\end{aligned}$$

Expanding the square results in the following required expectations,

$$E \left[\sum_{i \in \mathcal{D}} \left(\frac{n_o}{n} \right)^2 (\bar{Y}_{r2} - \bar{Y}_{a2o})^2 \right] = \left(\frac{n_o}{n} \right)^2 n_d \left[\sigma_{22} \left(\frac{1}{n_o} + \frac{1}{n} \right) + (\mu_{a2} - \mu_{r2})^2 \right],$$

$$E \left[\sum_{i \in \mathcal{D}} \left(\frac{n_o}{n} \right)^2 u_k^2 \right] = n_d \left(\frac{n_o}{n} \right)^2 \frac{\sigma_{2,1}}{n},$$

$$E \left[\sum_{i \in \mathcal{D}} \left(\frac{r}{q} + b_k \right)^2 (Y_{ai1} - \bar{Y}_{r1})^2 \right] = \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{2\sigma_{2,1}}{(n-1)\sigma_{11}} \right] n_d \left[\sigma_{11} \left(1 + \frac{1}{n} \right) + (\mu_{a1} - \mu_{r1})^2 \right],$$

$$E \left[\sum_{i \in \mathcal{D}} \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right)^2 (\bar{Y}_{a1d} - \bar{Y}_{r1})^2 \right]$$

$$\begin{aligned}
&= \binom{n_d}{n}^2 n_d \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{2\sigma_{2,1}}{(n-1)\sigma_{11}} \right] \left[\sigma_{11} \left(\frac{1}{n_d} + \frac{1}{n} \right) + (\mu_{a1} - \mu_{r1})^2 \right], \\
E \left[\sum_{i \in \mathcal{D}} \binom{n_d}{n}^2 \bar{e}_k^2 \right] &= \binom{n_d}{n}^2 \sigma_{2,1}, \\
E \left[\sum_{i \in \mathcal{D}} e_{i,k}^2 \right] &= n_d \sigma_{2,1}, \\
E \left[\sum_{i \in \mathcal{D}} 2 \frac{n_o}{n} (\bar{Y}_{r2} - \bar{Y}_{a2o}) \left(\frac{r}{q} + b_k \right) (Y_{ai1} - \bar{Y}_{r1}) \right] &= -2 \frac{n_o}{n} \frac{\sigma_{12}}{\sigma_{11}} n_d \left[\frac{\sigma_{12}}{n} + (\mu_{a1} - \mu_{r1})(\mu_{a2} - \mu_{r2}) \right], \\
E \left[\sum_{i \in \mathcal{D}} -2 \frac{n_d n_o}{n^2} (\bar{Y}_{r2} - \bar{Y}_{a2o}) \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{r1}) \right] &= 2 \frac{n_d^2 n_o}{n^2} \frac{\sigma_{12}}{\sigma_{11}} \left[\frac{\sigma_{12}}{n} + (\mu_{a1} - \mu_{r1})(\mu_{a2} - \mu_{r2}) \right], \\
E \left[\sum_{i \in \mathcal{D}} -2 \frac{n_d}{n} \left(\frac{r}{q} + b_k \right)^2 (\bar{Y}_{a1d} - \bar{Y}_{r1}) (Y_{ai1} - \bar{Y}_{r1}) \right] \\
&= -2 \frac{n_d}{n} \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{2\sigma_{2,1}}{(n-1)\sigma_{11}} \right] n_d \left[\sigma_{11} \left(\frac{1}{n_d} + \frac{1}{n} \right) + (\mu_{a1} - \mu_{r1})^2 \right], \\
E \left[\sum_{i \in \mathcal{D}} -2 \frac{n_d}{n} \bar{e}_k e_{i,k} \right] &= -2 \frac{n_d}{n} \sigma_{2,1}.
\end{aligned}$$

Combining the results over the observed and deviating cases for time 1 a baseline where $\mu_{a1} = \mu_{r1}$, $\pi_d = n_d/n$ and $\Delta = \mu_{a2} - \mu_{r2}$ results in (3.11). For $(1 + \frac{1}{K}) E(\hat{B})$ we require,

$$E[\hat{B}] = \frac{1}{K-1} E \left[\sum_{k=1}^K \left(\hat{\theta}_{\text{CR}, k} - \hat{\theta}_{\text{CR}, \text{MI}} \right)^2 \right],$$

where,

$$\hat{\theta}_{\text{CR}, k} = \frac{n_o}{n} \bar{Y}_{a2o} + \frac{n_d}{n} \left[\bar{Y}_{r2} + u_k + \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{r1}) + \bar{e}_k \right] - \bar{Y}_{r2},$$

$$\hat{\theta}_{\text{CR}, \text{MI}} = \frac{1}{K} \sum_{k=1}^K \hat{\theta}_{\text{CR}, k}.$$

That is,

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} u_k + \frac{n_d}{n} \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{r1}) + \frac{n_d}{n} \bar{e}_k - \frac{n_d}{n} \bar{u} - \frac{n_d}{n} \left(\frac{r}{q} + \bar{b} \right) (\bar{Y}_{a1d} - \bar{Y}_{r1}) - \frac{n_d}{n} \bar{e} \right)^2 \right].$$

Expanding the square and taking expectation by term gives the following non zero components,

$$\begin{aligned}
E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 u_k^2 \right] &= K \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{2,1}}{n}, \\
E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right)^2 (\bar{Y}_{a1d} - \bar{Y}_{r1})^2 \right] \\
&= K \left(\frac{n_d}{n} \right)^2 \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{2\sigma_{2,1}}{(n-1)\sigma_{11}} \right] \left[\sigma_{11} \left(\frac{1}{n_d} + \frac{1}{n} \right) + (\mu_{a1} - \mu_{r1})^2 \right], \\
E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \bar{e}_k^2 \right] &= K \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{2,1}}{n_d},
\end{aligned}$$

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \bar{u}^2 \right] = \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{2,1}}{n},$$

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + \bar{b} \right)^2 (\bar{Y}_{a1d} - \bar{Y}_{r1})^2 \right] \\ = K \left(\frac{n_d}{n} \right)^2 \left[\sigma_{11} \left(\frac{1}{n_d} + \frac{1}{n} \right) + (\mu_{a1} - \mu_{r1})^2 \right] \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{\sigma_{2,1}}{(n-1)\sigma_{11}} \left(\frac{K+1}{K} \right) \right],$$

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \bar{e}^2 \right] = \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{2,1}}{n_d},$$

$$E \left[\sum_{k=1}^K -2 \left(\frac{n_d}{n} \right)^2 u_k \bar{u} \right] = -2 \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{2,1}}{n},$$

$$E \left[\sum_{k=1}^K -2 \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right) \left(\frac{r}{q} + \bar{b} \right) (\bar{Y}_{a1d} - \bar{Y}_{r1})^2 \right] \\ = -2K \left(\frac{n_d}{n} \right)^2 \left[\sigma_{11} \left(\frac{1}{n_d} + \frac{1}{n} \right) + (\mu_{a1} - \mu_{r1})^2 \right] \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{\sigma_{2,1}}{(n-1)\sigma_{11}} \left(\frac{K+1}{K} \right) \right],$$

$$E \left[\sum_{k=1}^K -2 \left(\frac{n_d}{n} \right)^2 \bar{e}_k \bar{e} \right] = -2 \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{2,1}}{n_d}.$$

Combining the results for time 1 a baseline where $\mu_{a1} = \mu_{r1}$ and $\pi_d = n_d/n$ results in (3.12).

Proportion of Missing data	Derived Variance	Average Variance	MCSE
10% (N=25)	0.00506	0.00504	0.000011
20% (N=50)	0.00536	0.00537	0.000012
30% (N=75)	0.00568	0.00567	0.000014
40% (N=100)	0.00604	0.00606	0.000016
50% (N=125)	0.00644	0.00644	0.000019
60% (N=150)	0.00687	0.00687	0.000021
70% (N=175)	0.00733	0.00734	0.000025

Table B.3: Derived Rubin's variance versus simulated Rubin's variance (averaged over 1000 simulations) for the CR treatment effect in the asthma RCT inspired setting with $K=50$.

Proportion of Missing data	Derived Variance	Average Variance	MCSE
10% (N=25)	0.00426	0.00425	0.000009
20% (N=50)	0.00456	0.00457	0.000010
30% (N=75)	0.00488	0.00487	0.000013
40% (N=100)	0.00524	0.00527	0.000015
50% (N=125)	0.00564	0.00564	0.000018
60% (N=150)	0.00607	0.00607	0.000020
70% (N=175)	0.00653	0.00654	0.000024

Table B.4: Derived Rubin's variance versus simulated Rubin's variance (averaged over 1000 simulations) for the CR baseline adjusted treatment effect in the asthma RCT inspired setting with $K=50$.

B.4.3 J2R imputation calculations

Here we present additional details of the calculations involved in establishing the expectation of Rubin's variance estimator in the J2R case in Section 3.2.2. First we consider $E(\hat{W})$, which requires use to estimate under J2R,

$$(n-1)E(\hat{\sigma}_{22,a,k}) = E \left[\sum_{i \in \mathcal{O}} (Y_{ai2} - \hat{\mu}_{a2,k})^2 \right] + E \left[\sum_{i \in \mathcal{D}} (\hat{Y}_{ai2,k} - \hat{\mu}_{a2,k})^2 \right].$$

Decomposing the two-right hand components for the active arm separately, over the observed cases we have,

$$\begin{aligned} & E \left[\sum_{i \in \mathcal{O}} (Y_{ai2} - \hat{\mu}_{a2,k})^2 \right] \\ = & E \left[\sum_{i \in \mathcal{O}} \left(Y_{ai2} - \frac{n_o}{n} \bar{Y}_{a2o} - \frac{n_d}{n} \bar{Y}_{a2d,k} \right)^2 \right] \\ = & E \left[\sum_{i \in \mathcal{O}} \left(Y_{ai2} - \frac{n_o}{n} \bar{Y}_{a2o} - \frac{n_d}{n} \left[\bar{Y}_{r2} + u_k + \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{a1}) + \left(\frac{r}{q} + b_k \right) (r_k - a_k) + \bar{e}_k \right] \right)^2 \right] \\ = & E \left[\sum_{i \in \mathcal{O}} \left((Y_{ai2} - \bar{Y}_{a2o}) + \frac{n_d}{n} (\bar{Y}_{a2o} - \bar{Y}_{r2}) - \frac{n_d}{n} u_k - \frac{n_d}{n} \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{a1}) \right. \right. \\ & \left. \left. - \frac{n_d}{n} \left(\frac{r}{q} + b_k \right) (r_k - a_k) - \frac{n_d}{n} \bar{e}_k \right)^2 \right]. \end{aligned}$$

Expanding the square results in the following non-zero required expectations,

$$E \left[\sum_{i \in \mathcal{O}} (Y_{ai2} - \bar{Y}_{a2o})^2 \right] = (n_o - 1) \sigma_{22},$$

$$E \left[\sum_{i \in \mathcal{O}} \left(\frac{n_d}{n} \right)^2 (\bar{Y}_{a2o} - \bar{Y}_{r2})^2 \right] = \left(\frac{n_d}{n} \right)^2 \left[\sigma_{22} \left(1 + \frac{n_o}{n} \right) + n_o (\mu_{a2} - \mu_{r2})^2 \right],$$

$$E \left[\sum_{i \in \mathcal{O}} \left(\frac{n_d}{n} \right)^2 u_k^2 \right] = n_o \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{2,1}}{n},$$

$$E \left[\sum_{i \in \mathcal{O}} \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right)^2 (\bar{Y}_{a1d} - \bar{Y}_{a1})^2 \right] = \left(\frac{n_d}{n} \right)^2 n_o \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{2\sigma_{2,1}}{(n-1)\sigma_{11}} \right] \left[\sigma_{11} \left(\frac{1}{n_d} - \frac{1}{n} \right) \right],$$

$$E \left[\sum_{i \in \mathcal{O}} \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right)^2 (r_k - a_k)^2 \right] = \left(\frac{n_d}{n} \right)^2 n_o \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{2\sigma_{2,1}}{(n-1)\sigma_{11}} \right] \frac{2\sigma_{11}}{n},$$

$$E \left[\sum_{i \in \mathcal{O}} \left(\frac{n_d}{n} \right)^2 \bar{e}_k^2 \right] = \frac{n_d n_o}{n^2} \sigma_{2,1},$$

$$E \left[\sum_{i \in \mathcal{O}} -2 \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a2o} - \bar{Y}_{r2}) (\bar{Y}_{a1d} - \bar{Y}_{a1}) \right] = 2 \left(\frac{n_d}{n} \right)^2 \frac{n_o}{n} \frac{\sigma_{12}^2}{\sigma_{11}}.$$

Over the deviators we have,

$$\begin{aligned} & E \left[\sum_{i \in \mathcal{D}} (\hat{Y}_{ai2,k} - \hat{\mu}_{a2,k})^2 \right] \\ = & E \left[\sum_{i \in \mathcal{D}} \left(\bar{Y}_{r2} + u_k + \left(\frac{r}{q} + b_k \right) (Y_{ai1} - \bar{Y}_{a1}) + \left(\frac{r}{q} + b_k \right) (r_k - a_k) + e_{i,k} - \frac{n_o}{n} \bar{Y}_{a2o} - \frac{n_d}{n} \bar{Y}_{a2d,k} \right)^2 \right] \\ = & E \left[\sum_{i \in \mathcal{D}} \left(\frac{n_o}{n} (\bar{Y}_{r2} - \bar{Y}_{a2o}) + \frac{n_o}{n} u_k + \left(\frac{r}{q} + b_k \right) (Y_{ai1} - \bar{Y}_{a1}) \right. \right. \\ & \left. \left. - \frac{n_d}{n} \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{a1}) - \frac{n_d}{n} \bar{e}_k + e_{i,k} + \frac{n_o}{n} \left(\frac{r}{q} + b_k \right) (r_k - a_k) \right)^2 \right]. \end{aligned}$$

Expanding the square results in the following required expectations,

$$E \left[\sum_{i \in \mathcal{D}} \left(\frac{n_o}{n} \right)^2 (\bar{Y}_{r2} - \bar{Y}_{a2o})^2 \right] = \left(\frac{n_o}{n} \right)^2 n_d \left[\sigma_{22} \left(\frac{1}{n_o} + \frac{1}{n} \right) + (\mu_{a2} - \mu_{r2})^2 \right],$$

$$E \left[\sum_{i \in \mathcal{D}} \left(\frac{n_o}{n} \right)^2 u_k^2 \right] = n_d \left(\frac{n_o}{n} \right)^2 \frac{\sigma_{2,1}}{n},$$

$$E \left[\sum_{i \in \mathcal{D}} \left(\frac{r}{q} + b_k \right)^2 (Y_{ai1} - \bar{Y}_{a1})^2 \right] = \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{2\sigma_{2,1}}{(n-1)\sigma_{11}} \right] n_d \frac{\sigma_{11}(n-1)}{n},$$

$$E \left[\sum_{i \in \mathcal{D}} \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right)^2 (\bar{Y}_{a1d} - \bar{Y}_{a1})^2 \right] = \left(\frac{n_d}{n} \right)^2 n_d \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{2\sigma_{2,1}}{(n-1)\sigma_{11}} \right] \sigma_{11} \left[\frac{1}{n_d} - \frac{1}{n} \right],$$

$$E \left[\sum_{i \in \mathcal{D}} \left(\frac{n_d}{n} \right)^2 \bar{e}_k^2 \right] = \left(\frac{n_d}{n} \right)^2 \sigma_{2,1},$$

$$E \left[\sum_{i \in \mathcal{D}} e_{i,k}^2 \right] = n_d \sigma_{2,1},$$

$$E \left[\sum_{i \in \mathcal{D}} \left(\frac{n_o}{n} \right)^2 \left(\frac{r}{q} + b_k \right)^2 (r_k - a_k)^2 \right] = \left(\frac{n_o}{n} \right)^2 n_d \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{2\sigma_{2,1}}{(n-1)\sigma_{11}} \right] \frac{2\sigma_{11}}{n},$$

$$E \left[\sum_{i \in \mathcal{D}} 2 \frac{n_o}{n} (\bar{Y}_{r2} - \bar{Y}_{a2o}) \left(\frac{r}{q} + b_k \right) (Y_{ai1} - \bar{Y}_{a1}) \right] = 2 \frac{n_o}{n} \frac{n_d}{n} \frac{\sigma_{12}^2}{\sigma_{11}},$$

$$E \left[\sum_{i \in \mathcal{D}} -2 \frac{n_d n_o}{n^2} (\bar{Y}_{r2} - \bar{Y}_{a2o}) \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{a1}) \right] = -2 \left(\frac{n_d}{n} \right)^2 \left(\frac{n_o}{n} \right) \frac{\sigma_{12}^2}{\sigma_{11}},$$

$$E \left[\sum_{i \in \mathcal{D}} -2 \frac{n_d}{n} \left(\frac{r}{q} + b_k \right)^2 (\bar{Y}_{a1d} - \bar{Y}_{a1}) (Y_{ai1} - \bar{Y}_{a1}) \right] = -2 \frac{n_d^2}{n} \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{2\sigma_{2,1}}{(n-1)\sigma_{11}} \right] \sigma_{11} \left[\frac{1}{n_d} - \frac{1}{n} \right],$$

$$E \left[\sum_{i \in \mathcal{D}} -2 \frac{n_d}{n} \bar{e}_k e_{i,k} \right] = -2 \frac{n_d}{n} \sigma_{2,1}.$$

Combining the results over the observed and deviating cases for time 1 a baseline where $\mu_{a1} = \mu_{r1}$, $\pi_d = n_d/n$ and $\Delta = \mu_{a2} - \mu_{r2}$ results in (3.20). For $(1 + \frac{1}{K}) E(\hat{B})$ we require,

$$E[\hat{B}] = \frac{1}{K-1} E \left[\sum_{k=1}^K \left(\hat{\theta}_{J2R, k} - \hat{\theta}_{J2R, MI} \right)^2 \right],$$

where,

$$\hat{\theta}_{J2R, k} = \frac{n_o}{n} \bar{Y}_{a2o} + \frac{n_d}{n} \left[\bar{Y}_{r2} + u_k + \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{a1}) + \left(\frac{r}{q} + b_k \right) (r_k - a_k) + \bar{e}_k \right] - \bar{Y}_{r2},$$

$$\hat{\theta}_{J2R, MI} = \frac{1}{K} \sum_{k=1}^K \hat{\theta}_{J2R, k}.$$

That is,

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} u_k + \frac{n_d}{n} \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{a1}) + \frac{n_d}{n} \left(\frac{r}{q} + b_k \right) (r_k - a_k) + \frac{n_d}{n} \bar{e}_k - \frac{n_d}{n} \bar{u} - \frac{n_d}{n} \left(\frac{r}{q} + \bar{b} \right) (\bar{Y}_{a1d} - \bar{Y}_{a1}) - \frac{n_d}{n} \frac{r}{q} \bar{r} + \frac{n_d}{n} \frac{r}{q} \bar{a} - \frac{n_d}{n} \frac{1}{K} \sum_{i=1}^K b_k r_k + \frac{n_d}{n} \frac{1}{K} \sum_{i=1}^K b_k a_k - \frac{n_d}{n} \bar{e} \right)^2 \right].$$

Expanding the square and taking expectation by term gives the following non zero components,

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 u_k^2 \right] = K \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{2,1}}{n},$$

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right)^2 (\bar{Y}_{a1d} - \bar{Y}_{a1})^2 \right] = K \left(\frac{n_d}{n} \right)^2 \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{2\sigma_{2,1}}{(n-1)\sigma_{11}} \right] \sigma_{11} \left[\frac{1}{n_d} - \frac{1}{n} \right],$$

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \bar{e}_k^2 \right] = K \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{2,1}}{n_d},$$

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \bar{u}^2 \right] = \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{2,1}}{n},$$

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + \bar{b} \right)^2 (\bar{Y}_{a1d} - \bar{Y}_{a1})^2 \right] = K \left(\frac{n_d}{n} \right)^2 \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{\sigma_{2,1}}{(n-1)\sigma_{11}} \left(\frac{K+1}{K} \right) \right] \sigma_{11} \left[\frac{1}{n_d} - \frac{1}{n} \right],$$

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right)^2 (r_k - a_k)^2 \right] = K \left(\frac{n_d}{n} \right)^2 \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{2\sigma_{2,1}}{(n-1)\sigma_{11}} \right] \frac{2\sigma_{11}}{n},$$

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} \right)^2 \bar{r}^2 \right] = \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{11}}{n} \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{\sigma_{2,1}}{(n-1)\sigma_{11}} \right],$$

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} \right)^2 \bar{a}^2 \right] = \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{11}}{n} \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{\sigma_{2,1}}{(n-1)\sigma_{11}} \right],$$

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \left(\frac{1}{K} \right)^2 \left(\sum_{i=1}^K b_k r_k \right)^2 \right] = \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{2,1}}{n(n-1)},$$

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \left(\frac{1}{K} \right)^2 \left(\sum_{i=1}^K b_k a_k \right)^2 \right] = \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{2,1}}{n(n-1)},$$

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \bar{e}^2 \right] = \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{2,1}}{n_d},$$

$$E \left[\sum_{k=1}^K -2 \left(\frac{n_d}{n} \right)^2 u_k \bar{u} \right] = -2 \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{2,1}}{n},$$

$$E \left[\sum_{k=1}^K -2 \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right) \left(\frac{r}{q} + \bar{b} \right) (\bar{Y}_{a1d} - \bar{Y}_{a1})^2 \right] = -2K \left(\frac{n_d}{n} \right)^2 \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{\sigma_{2,1}}{(n-1)\sigma_{11}} \left(\frac{K+1}{K} \right) \right] \sigma_{11} \left[\frac{1}{n_d} - \frac{1}{n} \right],$$

$$E \left[\sum_{k=1}^K -2 \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right) (r_k - a_k) \frac{r}{q} \bar{r} \right] = -\frac{2\sigma_{11}}{n} \left(\frac{n_d}{n} \right)^2 \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{\sigma_{2,1}}{(n-1)\sigma_{11}} \right],$$

$$E \left[\sum_{k=1}^K 2 \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right) (r_k - a_k) \frac{r}{q} \bar{a} \right] = -\frac{2\sigma_{11}}{n} \left(\frac{n_d}{n} \right)^2 \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{\sigma_{2,1}}{(n-1)\sigma_{11}} \right],$$

$$E \left[\sum_{k=1}^K -2 \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right) (r_k - a_k) \frac{1}{K} \sum_{i=1}^K b_k r_k \right] = -2 \left(\frac{n_d}{n} \right)^2 \left[\frac{\sigma_{2,1}}{n(n-1)} \right],$$

$$E \left[\sum_{k=1}^K 2 \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right) (r_k - a_k) \frac{1}{K} \sum_{i=1}^K b_k a_k \right] = -2 \left(\frac{n_d}{n} \right)^2 \left[\frac{\sigma_{2,1}}{n(n-1)} \right],$$

$$E \left[\sum_{k=1}^K -2 \left(\frac{n_d}{n} \right)^2 \bar{e}_k \bar{e} \right] = -2 \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{2,1}}{n_d}.$$

Combining the results for time 1 a baseline where $\mu_{a1} = \mu_{r1}$ and $\pi_d = n_d/n$ results in (3.21).

Proportion of Missing data	Derived Variance	Average Variance	MCSE
10% (N=25)	0.00507	0.00505	0.000011
20% (N=50)	0.00539	0.00540	0.000012
30% (N=75)	0.00576	0.00575	0.000014
40% (N=100)	0.00618	0.00621	0.000017
50% (N=125)	0.00665	0.00666	0.000020
60% (N=150)	0.00717	0.00717	0.000024
70% (N=175)	0.00774	0.00776	0.000028

Table B.5: Derived Rubin’s variance versus simulated Rubin’s variance (averaged over 1000 simulations) for the J2R treatment effect in the asthma RCT inspired setting with $K=50$.

Proportion of Missing data	Derived Variance	Average Variance	MCSE
10% (N=25)	0.00427	0.00426	0.000009
20% (N=50)	0.00459	0.00460	0.000011
30% (N=75)	0.00496	0.00495	0.000013
40% (N=100)	0.00538	0.00541	0.000015
50% (N=125)	0.00585	0.00586	0.000019
60% (N=150)	0.00637	0.00637	0.000022
70% (N=175)	0.00694	0.00697	0.000026

Table B.6: Derived Rubin’s variance versus simulated Rubin’s variance (averaged over 1000 simulations) for the J2R baseline adjusted treatment effect in the asthma RCT inspired setting with $K=50$.

B.4.4 CIR imputation calculations

Here we present additional details of the calculations involved in establishing the expectation of Rubin’s variance estimator in the CIR case in Section 3.2.3. First we consider $E(\hat{W})$, which requires use to estimate under CIR,

$$(n-1) E(\hat{\sigma}_{22,a,k}) = E \left[\sum_{i \in \mathcal{O}} (Y_{ai2} - \hat{\mu}_{a2,k})^2 \right] + E \left[\sum_{i \in \mathcal{D}} (\hat{Y}_{ai2,k} - \hat{\mu}_{a2,k})^2 \right].$$

Decomposing the two-right hand components for the active arm separately, over the observed cases we have,

$$\begin{aligned} & E \left[\sum_{i \in \mathcal{O}} (Y_{ai2} - \hat{\mu}_{a2,k})^2 \right] \\ = & E \left[\sum_{i \in \mathcal{O}} \left(Y_{ai2} - \frac{n_o}{n} \bar{Y}_{a2o} - \frac{n_d}{n} \bar{Y}_{a2d,k} \right)^2 \right] \end{aligned}$$

$$\begin{aligned}
&= E \left[\sum_{i \in \mathcal{O}} \left(Y_{ai2} - \frac{n_o}{n} \bar{Y}_{a2o} - \frac{n_d}{n} \left[\bar{Y}_{a1} + \bar{Y}_{r2} - \bar{Y}_{r1} + u_k + \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{a1}) \right. \right. \right. \\
&\quad \left. \left. \left. + \left(\frac{r}{q} + b_k \right) (r_k - a_k) + (a_k - r_k) + \bar{e}_k \right] \right)^2 \right] \\
&= E \left[\sum_{i \in \mathcal{O}} \left((Y_{ai2} - \bar{Y}_{a2o}) + \frac{n_d}{n} (\bar{Y}_{a2o} - \bar{Y}_{r2}) + \frac{n_d}{n} (\bar{Y}_{r1} - \bar{Y}_{a1}) - \frac{n_d}{n} u_k - \frac{n_d}{n} \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{a1}) \right. \right. \\
&\quad \left. \left. - \frac{n_d}{n} \left(\frac{r}{q} + b_k \right) (r_k - a_k) - \frac{n_d}{n} (a_k - r_k) - \frac{n_d}{n} \bar{e}_k \right)^2 \right].
\end{aligned}$$

Expanding the square results in the same non-zero expectations as under J2R, plus the following non-zero expectations,

$$E \left[\sum_{i \in \mathcal{O}} \left(\frac{n_d}{n} \right)^2 (\bar{Y}_{r1} - \bar{Y}_{a1})^2 \right] = 2 \left(\frac{n_d}{n} \right)^2 \frac{n_o}{n} \sigma_{11},$$

$$E \left[\sum_{i \in \mathcal{O}} \left(\frac{n_d}{n} \right)^2 (a_k - r_k)^2 \right] = 2 \left(\frac{n_d}{n} \right)^2 \frac{n_o}{n} \sigma_{11},$$

$$E \left[\sum_{i \in \mathcal{O}} 2 \left(\frac{n_d}{n} \right)^2 (\bar{Y}_{r1} - \bar{Y}_{a1}) (\bar{Y}_{a2o} - \bar{Y}_{r2}) \right] = -4 \left(\frac{n_d}{n} \right)^2 \frac{n_o}{n} \sigma_{12},$$

$$E \left[\sum_{i \in \mathcal{O}} 2 \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right) (r_k - a_k) (a_k - r_k) \right] = -4 \left(\frac{n_d}{n} \right)^2 \frac{n_o}{n} \sigma_{12}.$$

Over the deviators we have,

$$\begin{aligned}
&E \left[\sum_{i \in \mathcal{D}} \left(\hat{Y}_{ai2,k} - \hat{\mu}_{a2,k} \right)^2 \right] \\
&= E \left[\sum_{i \in \mathcal{D}} \left(\bar{Y}_{a1} + \bar{Y}_{r2} - \bar{Y}_{r1} + u_k + \left(\frac{r}{q} + b_k \right) (Y_{ai1} - \bar{Y}_{a1}) + \left(\frac{r}{q} + b_k \right) (r_k - a_k) + a_k - r_k + e_{i,k} \right. \right. \\
&\quad \left. \left. - \frac{n_o}{n} \bar{Y}_{a2o} - \frac{n_d}{n} \bar{Y}_{a2d,k} \right)^2 \right] \\
&= E \left[\sum_{i \in \mathcal{D}} \left(\frac{n_o}{n} (\bar{Y}_{r2} - \bar{Y}_{a2o}) + \frac{n_o}{n} (\bar{Y}_{a1} - \bar{Y}_{r1}) + \frac{n_o}{n} u_k + \left(\frac{r}{q} + b_k \right) (Y_{ai1} - \bar{Y}_{a1}) \right. \right. \\
&\quad \left. \left. - \frac{n_d}{n} \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{a1}) - \frac{n_d}{n} \bar{e}_k + e_{i,k} + \frac{n_o}{n} \left(\frac{r}{q} + b_k \right) (r_k - a_k) + \frac{n_o}{n} (a_k - r_k) \right)^2 \right].
\end{aligned}$$

Expanding the square results in the same non-zero expectations as under J2R, plus the following non-zero expectations,

$$E \left[\sum_{i \in \mathcal{D}} \left(\frac{n_o}{n} \right)^2 (\bar{Y}_{a1} - \bar{Y}_{r1})^2 \right] = \left(\frac{n_o}{n} \right)^2 n_d \frac{2\sigma_{11}}{n},$$

$$E \left[\sum_{i \in \mathcal{D}} \left(\frac{n_o}{n} \right)^2 (a_k - r_k)^2 \right] = \left(\frac{n_o}{n} \right)^2 n_d \frac{2\sigma_{11}}{n},$$

$$E \left[\sum_{i \in \mathcal{D}} 2 \left(\frac{n_o}{n} \right)^2 (\bar{Y}_{a1} - \bar{Y}_{r1}) (\bar{Y}_{r2} - \bar{Y}_{a2o}) \right] = -4 \left(\frac{n_o}{n} \right)^2 n_d \frac{\sigma_{12}}{n},$$

$$E \left[\sum_{i \in \mathcal{D}} 2 \left(\frac{n_o}{n} \right)^2 \left(\frac{r}{q} + b_k \right) (r_k - a_k) (a_k - r_k) \right] = -4 \left(\frac{n_o}{n} \right)^2 n_d \frac{\sigma_{12}}{n}.$$

Combining the results over the observed and deviating cases for time 1 a baseline where $\mu_{a1} = \mu_{r1}$, $\pi_d = n_d/n$ and $\Delta = \mu_{a2} - \mu_{r2}$ results in (3.27). For $(1 + \frac{1}{K}) E(\hat{B})$ we require,

$$E[\hat{B}] = \frac{1}{K-1} E \left[\sum_{k=1}^K \left(\hat{\theta}_{\text{CIR}, k} - \hat{\theta}_{\text{CIR}, \text{MI}} \right)^2 \right],$$

where,

$$\begin{aligned}\hat{\theta}_{\text{CIR}, k} &= \frac{n_o}{n} \bar{Y}_{a2o} + \frac{n_d}{n} \left[\bar{Y}_{a1} + \bar{Y}_{r2} - \bar{Y}_{r1} + u_k + \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{a1}) + \left(\frac{r}{q} + b_k \right) (r_k - a_k) \right. \\ &\quad \left. + (a_k - r_k) + \bar{e}_k \right] - \bar{Y}_{r2}, \\ \hat{\theta}_{\text{CIR}, \text{MI}} &= \frac{1}{K} \sum_{k=1}^K \hat{\theta}_{\text{CIR}, k}.\end{aligned}$$

That is,

$$\begin{aligned}E \left[\sum_{k=1}^K \left(\frac{n_d}{n} u_k + \frac{n_d}{n} \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{a1}) + \frac{n_d}{n} \left(\frac{r}{q} + b_k \right) (r_k - a_k) + \frac{n_d}{n} \bar{e}_k - \frac{n_d}{n} \bar{u} \right. \right. \\ \left. \left. - \frac{n_d}{n} \left(\frac{r}{q} + \bar{b} \right) (\bar{Y}_{a1d} - \bar{Y}_{a1}) - \frac{n_d}{n} \frac{r}{q} \bar{r} + \frac{n_d}{n} \frac{r}{q} \bar{a} - \frac{n_d}{n} \frac{1}{K} \sum_{i=1}^K b_k r_k + \frac{n_d}{n} \frac{1}{K} \sum_{i=1}^K b_k a_k - \frac{n_d}{n} \bar{e} \right. \right. \\ \left. \left. + \frac{n_d}{n} (a_k - r_k) - \frac{n_d}{n} \bar{a} + \frac{n_d}{n} \bar{r} \right)^2 \right].\end{aligned}$$

Expanding the square and taking expectation by term results in the same non-zero expectations as under J2R, plus the following non-zero expectations,

$$\begin{aligned}E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 (a_k - r_k)^2 \right] &= K \left(\frac{n_d}{n} \right)^2 \frac{2\sigma_{11}}{n}, \\ E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \bar{a}^2 \right] &= \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{11}}{n}, \\ E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \bar{r}^2 \right] &= \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{11}}{n}, \\ E \left[\sum_{k=1}^K 2 \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right) (r_k - a_k) (a_k - r_k) \right] &= -4K \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{12}}{n}, \\ E \left[\sum_{k=1}^K -2 \left(\frac{n_d}{n} \right)^2 \frac{r}{q} \bar{r} (a_k - r_k) \right] &= 2 \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{12}}{n}, \\ E \left[\sum_{k=1}^K 2 \left(\frac{n_d}{n} \right)^2 \frac{r}{q} \bar{a} (a_k - r_k) \right] &= 2 \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{12}}{n}, \\ E \left[\sum_{k=1}^K -2 \left(\frac{n_d}{n} \right)^2 \bar{a} (a_k - r_k) \right] &= -2 \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{11}}{n}, \\ E \left[\sum_{k=1}^K 2 \left(\frac{n_d}{n} \right)^2 \bar{r} (a_k - r_k) \right] &= -2 \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{11}}{n}, \\ E \left[\sum_{k=1}^K -2 \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right) \bar{a} (r_k - a_k) \right] &= 2 \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{12}}{n}, \\ E \left[\sum_{k=1}^K -2 \left(\frac{n_d}{n} \right)^2 \frac{r}{q} \bar{a}^2 \right] &= -2 \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{12}}{n}, \\ E \left[\sum_{k=1}^K 2 \left(\frac{n_d}{n} \right)^2 \bar{r} \left(\frac{r}{q} + b_k \right) (r_k - a_k) \right] &= 2 \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{12}}{n}, \\ E \left[\sum_{k=1}^K -2 \left(\frac{n_d}{n} \right)^2 \frac{r}{q} \bar{r}^2 \right] &= -2 \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{12}}{n}.\end{aligned}$$

Combining the results for time 1 a baseline where $\mu_{a1} = \mu_{r1}$ and $\pi_d = n_d/n$ results in (3.28).

Proportion of Missing data	Derived Variance	Average Variance	MCSE
10% (N=25)	0.00507	0.00505	0.000011
20% (N=50)	0.00539	0.00540	0.000012
30% (N=75)	0.00576	0.00575	0.000014
40% (N=100)	0.00618	0.00620	0.000016
50% (N=125)	0.00665	0.00665	0.000019
60% (N=150)	0.00717	0.00717	0.000023
70% (N=175)	0.00774	0.00776	0.000026

Table B.7: Derived Rubin’s variance versus simulated Rubin’s variance (averaged over 1000 simulations) for the CIR treatment effect in the asthma RCT inspired setting with $K=50$.

Proportion of Missing data	Derived Variance	Average Variance	MCSE
10% (N=25)	0.00427	0.00426	0.000009
20% (N=50)	0.00459	0.00460	0.000011
30% (N=75)	0.00496	0.00495	0.000013
40% (N=100)	0.00538	0.00540	0.000016
50% (N=125)	0.00585	0.00586	0.000019
60% (N=150)	0.00637	0.00637	0.000022
70% (N=175)	0.00694	0.00697	0.000026

Table B.8: Derived Rubin’s variance versus simulated Rubin’s variance (averaged over 1000 simulations) for the CIR baseline adjusted treatment effect in the asthma RCT inspired setting with $K=50$.

B.4.5 LMCF imputation calculations

Here we present additional details of the calculations involved in establishing the expectation of Rubin’s variance estimator in the LMCF case in Section 3.2.4. First we consider $E(\hat{W})$, which requires use to estimate under LMCF,

$$(n-1)E(\hat{\sigma}_{22,a,k}) = E \left[\sum_{i \in \mathcal{O}} (Y_{ai2} - \hat{\mu}_{a2,k})^2 \right] + E \left[\sum_{i \in \mathcal{D}} (\hat{Y}_{ai2,k} - \hat{\mu}_{a2,k})^2 \right].$$

Decomposing the two-right hand components for the active arm separately, over the observed cases we have,

$$\begin{aligned} & E \left[\sum_{i \in \mathcal{O}} (Y_{ai2} - \hat{\mu}_{a2,k})^2 \right] \\ &= E \left[\sum_{i \in \mathcal{O}} \left(Y_{ai2} - \frac{n_o}{n} \bar{Y}_{a2o} - \frac{n_d}{n} \bar{Y}_{a2d,k} \right)^2 \right] \\ &= E \left[\sum_{i \in \mathcal{O}} \left(Y_{ai2} - \frac{n_o}{n} \bar{Y}_{a2o} - \frac{n_d}{n} \left[\bar{Y}_{a1} + u_k + \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{a1} - u_k) + \bar{e}_k \right] \right)^2 \right] \\ &= E \left[\sum_{i \in \mathcal{O}} \left((Y_{ai2} - \bar{Y}_{a2o}) + \frac{n_d}{n} (\bar{Y}_{a2o} - \bar{Y}_{a1}) - \frac{n_d}{n} u_k - \frac{n_d}{n} \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{a1}) \right)^2 \right] \end{aligned}$$

$$+ \frac{n_d}{n} u_k \left(\frac{r}{q} + b_k \right) - \frac{n_d}{n} \bar{e}_k \right)^2 \Big].$$

Expanding the square results in the following non-zero required expectations,

$$E \left[\sum_{i \in \mathcal{O}} (Y_{ai2} - \bar{Y}_{a2o})^2 \right] = (n_o - 1) \sigma_{22},$$

$$E \left[\sum_{i \in \mathcal{O}} \left(\frac{n_d}{n} \right)^2 (\bar{Y}_{a2o} - \bar{Y}_{a1})^2 \right] = \left(\frac{n_d}{n} \right)^2 n_o \left[(\mu_{a2} - \mu_{a1})^2 + \frac{\sigma_{22}}{n_o} + \frac{\sigma_{11} - 2\sigma_{12}}{n} \right],$$

$$E \left[\sum_{i \in \mathcal{O}} \left(\frac{n_d}{n} \right)^2 u_k^2 \right] = n_o \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{11}}{n},$$

$$E \left[\sum_{i \in \mathcal{O}} \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right)^2 (\bar{Y}_{a1d} - \bar{Y}_{a1})^2 \right] = \left(\frac{n_d}{n} \right)^2 n_o \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{2\sigma_{2,1}}{(n_o - 1)\sigma_{11}} \right] \left[\sigma_{11} \left(\frac{1}{n_d} - \frac{1}{n} \right) \right],$$

$$E \left[\sum_{i \in \mathcal{O}} \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right)^2 u_k^2 \right] = \left(\frac{n_d}{n} \right)^2 n_o \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{2\sigma_{2,1}}{(n_o - 1)\sigma_{11}} \right] \frac{\sigma_{11}}{n},$$

$$E \left[\sum_{i \in \mathcal{O}} \left(\frac{n_d}{n} \right)^2 \bar{e}_k^2 \right] = \frac{n_d n_o}{n^2} \sigma_{2,1},$$

$$E \left[\sum_{i \in \mathcal{O}} -2 \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a2o} - \bar{Y}_{a1}) (\bar{Y}_{a1d} - \bar{Y}_{a1}) \right] = 2 \left(\frac{n_d}{n} \right)^2 \frac{n_o}{n} \frac{\sigma_{12}^2}{\sigma_{11}},$$

$$E \left[\sum_{i \in \mathcal{O}} -2 \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right) u_k^2 \right] = -2 \left(\frac{n_d}{n} \right)^2 \frac{n_o}{n} \sigma_{12}.$$

Over the deviators we have,

$$\begin{aligned} & E \left[\sum_{i \in \mathcal{D}} \left(\hat{Y}_{ai2,k} - \hat{\mu}_{a2,k} \right)^2 \right] \\ &= E \left[\sum_{i \in \mathcal{D}} \left(\bar{Y}_{a1} + u_k + \left(\frac{r}{q} + b_k \right) (Y_{ai1} - \bar{Y}_{a1}) - \left(\frac{r}{q} + b_k \right) u_k + e_{i,k} - \frac{n_o}{n} \bar{Y}_{a2o} - \frac{n_d}{n} \bar{Y}_{a2d,k} \right)^2 \right] \\ &= E \left[\sum_{i \in \mathcal{D}} \left(\frac{n_o}{n} (\bar{Y}_{a1} - \bar{Y}_{a2o}) + \frac{n_o}{n} u_k + \left(\frac{r}{q} + b_k \right) (Y_{ai1} - \bar{Y}_{a1}) \right. \right. \\ &\quad \left. \left. - \frac{n_d}{n} \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{a1}) - \frac{n_d}{n} \bar{e}_k + e_{i,k} - \frac{n_o}{n} \left(\frac{r}{q} + b_k \right) u_k \right)^2 \right]. \end{aligned}$$

Expanding the square results in the following required expectations,

$$E \left[\sum_{i \in \mathcal{D}} \left(\frac{n_o}{n} \right)^2 (\bar{Y}_{a1} - \bar{Y}_{a2o})^2 \right] = \left(\frac{n_o}{n} \right)^2 n_d \left[(\mu_{a2} - \mu_{a1})^2 + \frac{\sigma_{22}}{n_o} + \frac{\sigma_{11} - 2\sigma_{12}}{n} \right],$$

$$E \left[\sum_{i \in \mathcal{D}} \left(\frac{n_o}{n} \right)^2 u_k^2 \right] = n_d \left(\frac{n_o}{n} \right)^2 \frac{\sigma_{11}}{n},$$

$$E \left[\sum_{i \in \mathcal{D}} \left(\frac{r}{q} + b_k \right)^2 (Y_{ai1} - \bar{Y}_{a1})^2 \right] = \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{2\sigma_{2,1}}{(n_o - 1)\sigma_{11}} \right] n_d \frac{\sigma_{11}(n-1)}{n},$$

$$E \left[\sum_{i \in \mathcal{D}} \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right)^2 (\bar{Y}_{a1d} - \bar{Y}_{a1})^2 \right] = \left(\frac{n_d}{n} \right)^2 n_d \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{2\sigma_{2,1}}{(n_o - 1)\sigma_{11}} \right] \sigma_{11} \left[\frac{1}{n_d} - \frac{1}{n} \right],$$

$$E \left[\sum_{i \in \mathcal{D}} \left(\frac{n_d}{n} \right)^2 \bar{e}_k^2 \right] = \left(\frac{n_d}{n} \right)^2 \sigma_{2,1},$$

$$E \left[\sum_{i \in \mathcal{D}} e_{i,k}^2 \right] = n_d \sigma_{2,1},$$

$$E \left[\sum_{i \in \mathcal{D}} \left(\frac{n_o}{n} \right)^2 \left(\frac{r}{q} + b_k \right)^2 u_k^2 \right] = \left(\frac{n_o}{n} \right)^2 n_d \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{2\sigma_{2,1}}{(n_o-1)\sigma_{11}} \right] \frac{\sigma_{11}}{n},$$

$$E \left[\sum_{i \in \mathcal{D}} 2 \frac{n_o}{n} (\bar{Y}_{a1} - \bar{Y}_{a2o}) \left(\frac{r}{q} + b_k \right) (Y_{ai1} - \bar{Y}_{a1}) \right] = 2 \frac{n_o}{n} \frac{n_d}{n} \frac{\sigma_{12}^2}{\sigma_{11}},$$

$$E \left[\sum_{i \in \mathcal{D}} -2 \frac{n_d n_o}{n^2} (\bar{Y}_{a1} - \bar{Y}_{a2o}) \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{a1}) \right] = -2 \left(\frac{n_d}{n} \right)^2 \left(\frac{n_o}{n} \right) \frac{\sigma_{12}^2}{\sigma_{11}},$$

$$E \left[\sum_{i \in \mathcal{D}} -2 \left(\frac{n_o}{n} \right)^2 u_k^2 \left(\frac{r}{q} + b_k \right) \right] = -2 \left(\frac{n_o}{n} \right)^2 \left(\frac{n_d}{n} \right) \sigma_{12},$$

$$\begin{aligned} E \left[\sum_{i \in \mathcal{D}} -2 \frac{n_d}{n} \left(\frac{r}{q} + b_k \right)^2 (\bar{Y}_{a1d} - \bar{Y}_{a1}) (Y_{ai1} - \bar{Y}_{a1}) \right] \\ = -2 \frac{n_d}{n} \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{2\sigma_{2,1}}{(n_o-1)\sigma_{11}} \right] \sigma_{11} \left[\frac{1}{n_d} - \frac{1}{n} \right], \end{aligned}$$

$$E \left[\sum_{i \in \mathcal{D}} -2 \frac{n_d}{n} \bar{e}_k e_{i,k} \right] = -2 \frac{n_d}{n} \sigma_{2,1}.$$

Combining the results over the observed and deviating cases for time 1 a baseline where $\mu_{a1} = \mu_{r1}$, $\pi_d = n_d/n$ and $\Delta = \mu_{a2} - \mu_{a1}$ results in 3.34. For $(1 + \frac{1}{K}) E(\hat{B})$ we require,

$$E[\hat{B}] = \frac{1}{K-1} E \left[\sum_{k=1}^K \left(\hat{\theta}_{\text{LMCF}, k} - \hat{\theta}_{\text{LMCF}, \text{MI}} \right)^2 \right],$$

where,

$$\hat{\theta}_{\text{LMCF}, k} = \frac{n_o}{n} \bar{Y}_{a2o} + \frac{n_d}{n} \left[\bar{Y}_{a1} + u_k + \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{a1}) - \left(\frac{r}{q} + b_k \right) u_k + \bar{e}_k \right] - \bar{Y}_{r2},$$

$$\hat{\theta}_{\text{LMCF}, \text{MI}} = \frac{1}{K} \sum_{k=1}^K \hat{\theta}_{\text{LMCF}, k}.$$

That is,

$$\begin{aligned} E \left[\sum_{k=1}^K \left(\frac{n_d}{n} u_k + \frac{n_d}{n} \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{a1}) - \frac{n_d}{n} \left(\frac{r}{q} + b_k \right) u_k + \frac{n_d}{n} \bar{e}_k - \frac{n_d}{n} \bar{u} \right. \right. \\ \left. \left. - \frac{n_d}{n} \left(\frac{r}{q} + \bar{b} \right) (\bar{Y}_{a1d} - \bar{Y}_{a1}) + \frac{n_d}{n} \frac{r}{q} \bar{u} + \frac{n_d}{n} \frac{1}{K} \sum_{i=1}^K b_k u_k - \frac{n_d}{n} \bar{e} \right)^2 \right]. \end{aligned}$$

Expanding the square and taking expectation by term gives the following non zero components,

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 u_k^2 \right] = K \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{11}}{n},$$

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right)^2 (\bar{Y}_{a1d} - \bar{Y}_{a1})^2 \right] = K \left(\frac{n_d}{n} \right)^2 \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{2\sigma_{2,1}}{(n_o-1)\sigma_{11}} \right] \sigma_{11} \left[\frac{1}{n_d} - \frac{1}{n} \right],$$

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \bar{e}_k^2 \right] = K \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{2,1}}{n_d},$$

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \bar{u}^2 \right] = \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{11}}{n},$$

$$\begin{aligned}
E & \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + \bar{b} \right)^2 (\bar{Y}_{a1d} - \bar{Y}_{a1})^2 \right] \\
& = K \left(\frac{n_d}{n} \right)^2 \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{\sigma_{2,1}}{(n_o-1)\sigma_{11}} \left(\frac{K+1}{K} \right) \right] \sigma_{11} \left[\frac{1}{n_d} - \frac{1}{n} \right], \\
E & \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right)^2 u_k^2 \right] = K \left(\frac{n_d}{n} \right)^2 \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{2\sigma_{2,1}}{(n_o-1)\sigma_{11}} \right] \frac{\sigma_{11}}{n}, \\
E & \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} \right)^2 \bar{u}^2 \right] = \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{11}}{n} \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{2\sigma_{2,1}}{(n_o-1)\sigma_{11}} \right], \\
E & \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \left(\frac{1}{K} \right)^2 \left(\sum_{i=1}^K b_k u_k \right)^2 \right] = \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{2,1}}{n(n_o-1)}, \\
E & \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \bar{e}^2 \right] = \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{2,1}}{n_d}, \\
E & \left[\sum_{k=1}^K -2 \left(\frac{n_d}{n} \right)^2 u_k \bar{u} \right] = -2 \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{11}}{n}, \\
E & \left[\sum_{k=1}^K -2 \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right) \left(\frac{r}{q} + \bar{b} \right) (\bar{Y}_{a1d} - \bar{Y}_{a1})^2 \right] \\
& = -2K \left(\frac{n_d}{n} \right)^2 \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{\sigma_{2,1}}{(n_o-1)\sigma_{11}} \left(\frac{K+1}{K} \right) \right] \sigma_{11} \left[\frac{1}{n_d} - \frac{1}{n} \right], \\
E & \left[\sum_{k=1}^K -2 \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right) u_k^2 \right] = -2K \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{12}}{n}, \\
E & \left[\sum_{k=1}^K 2 \left(\frac{n_d}{n} \right)^2 \frac{r}{q} u_k \bar{u} \right] = 2 \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{12}}{n}, \\
E & \left[\sum_{k=1}^K 2 \left(\frac{n_d}{n} \right)^2 \left(\frac{r}{q} + b_k \right) u_k \bar{u} \right] = 2 \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{12}}{n}, \\
E & \left[\sum_{k=1}^K -2 \left(\frac{n_d}{n} \right)^2 \frac{r}{q} \bar{u}^2 \right] = -2 \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{12}}{n}, \\
E & \left[\sum_{k=1}^K -2 \left(\frac{n_d}{n} \right)^2 \frac{r}{q} \bar{u} u_k \left(\frac{r}{q} + b_k \right) \right] = -2 \left(\frac{n_d}{n} \right)^2 \left[\left(\frac{\sigma_{12}}{\sigma_{11}} \right)^2 + \frac{2\sigma_{2,1}}{(n_o-1)\sigma_{11}} \right] \frac{\sigma_{11}}{n}, \\
E & \left[\sum_{k=1}^K -2 \left(\frac{n_d}{n} \right)^2 u_k \left(\frac{r}{q} + b_k \right) \frac{1}{K} \sum_{i=1}^K u_k b_k \right] = -2 \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{2,1}}{n(n_o-1)}, \\
E & \left[\sum_{k=1}^K -2 \left(\frac{n_d}{n} \right)^2 \bar{e}_k \bar{e} \right] = -2 \left(\frac{n_d}{n} \right)^2 \frac{\sigma_{2,1}}{n_d}.
\end{aligned}$$

Combining the results for time 1 a baseline where $\mu_{a1} = \mu_{r1}$ and $\pi_d = n_d/n$ results in (3.35).

Proportion of Missing data	Derived Variance	Average Variance	MCSE
10% (N=25)	0.00503	0.00500	0.000010
20% (N=50)	0.00526	0.00528	0.000011
30% (N=75)	0.00549	0.00549	0.000013
40% (N=100)	0.00574	0.00574	0.000014
50% (N=125)	0.00599	0.00601	0.000017
60% (N=150)	0.00625	0.00627	0.000020
70% (N=175)	0.00653	0.00656	0.000023

Table B.9: Derived Rubin's variance versus simulated Rubin's variance under LMCF
Derived Rubin's variance versus simulated Rubin's variance (averaged over 1000 simulations) for
the LMCF treatment effect in the asthma RCT inspired setting with $K=50$.

Proportion of Missing data	Derived Variance	Average Variance	MCSE
10% (N=25)	0.00422	0.00421	0.000009
20% (N=50)	0.00446	0.00448	0.000010
30% (N=75)	0.00469	0.00469	0.000012
40% (N=100)	0.00493	0.00494	0.000013
50% (N=125)	0.00518	0.00522	0.000016
60% (N=150)	0.00544	0.00547	0.000019
70% (N=175)	0.00571	0.00576	0.000023

Table B.10: Derived Rubin's variance versus simulated Rubin's variance (averaged over 1000
simulations) for the LMCF baseline adjusted treatment effect in the asthma RCT inspired setting
with $K=50$.

Appendix C

Additional results from baseline and single follow-up simulation studies

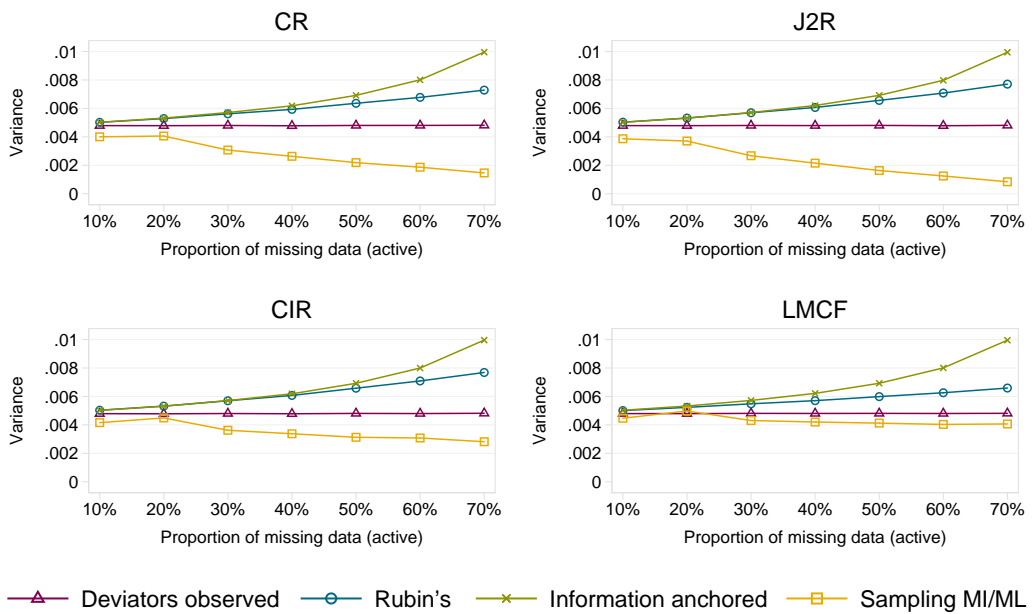


Figure C.1: Rubin's variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with $\Delta = 0$ and $n=250$ per arm.

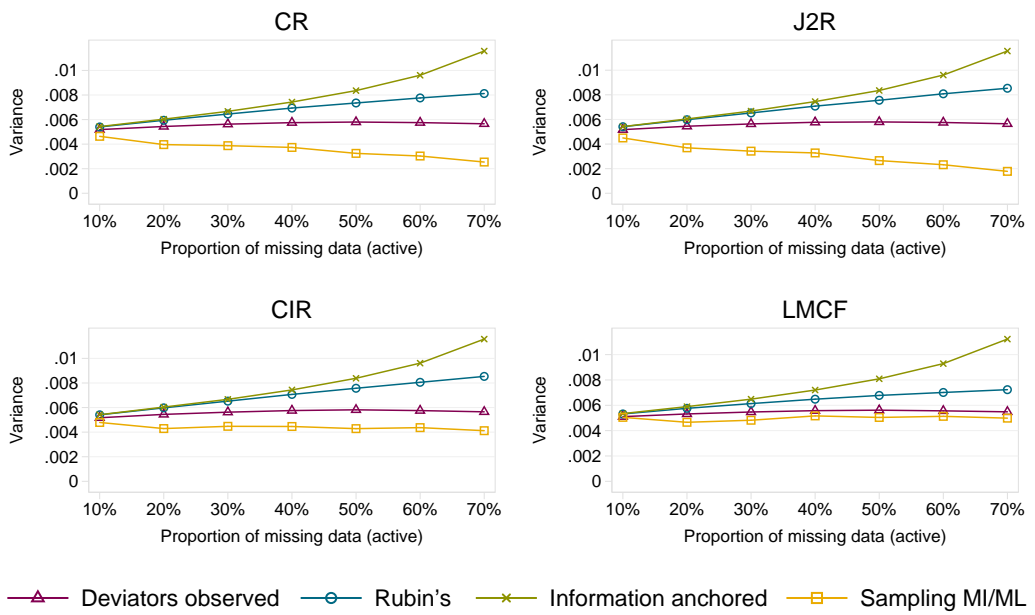


Figure C.2: Rubin's variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with $\Delta = 1$ and $n=250$ per arm.

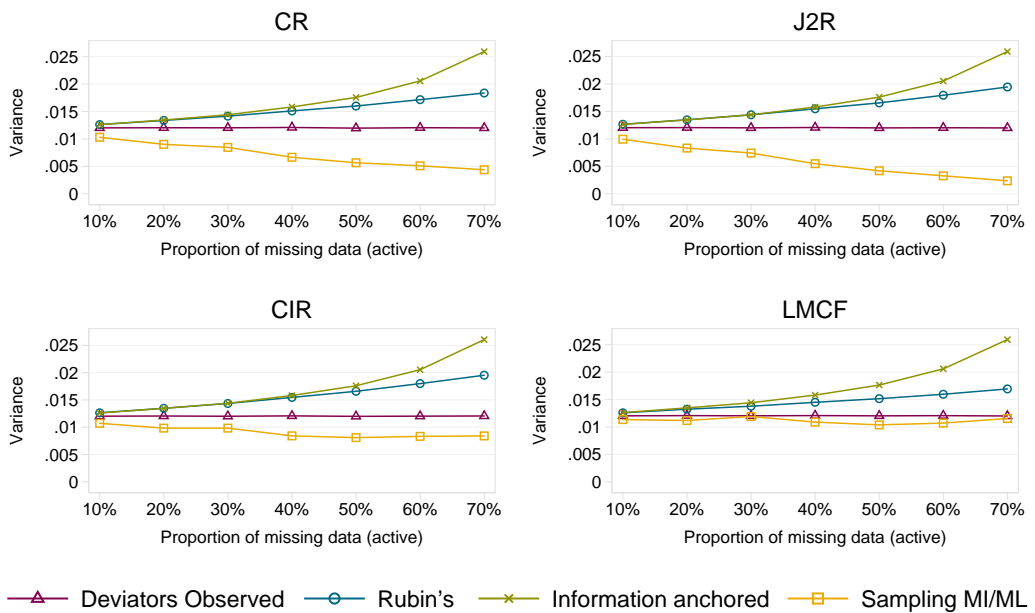


Figure C.3: Rubin's variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with $\Delta = 0$ and $n=100$ per arm.

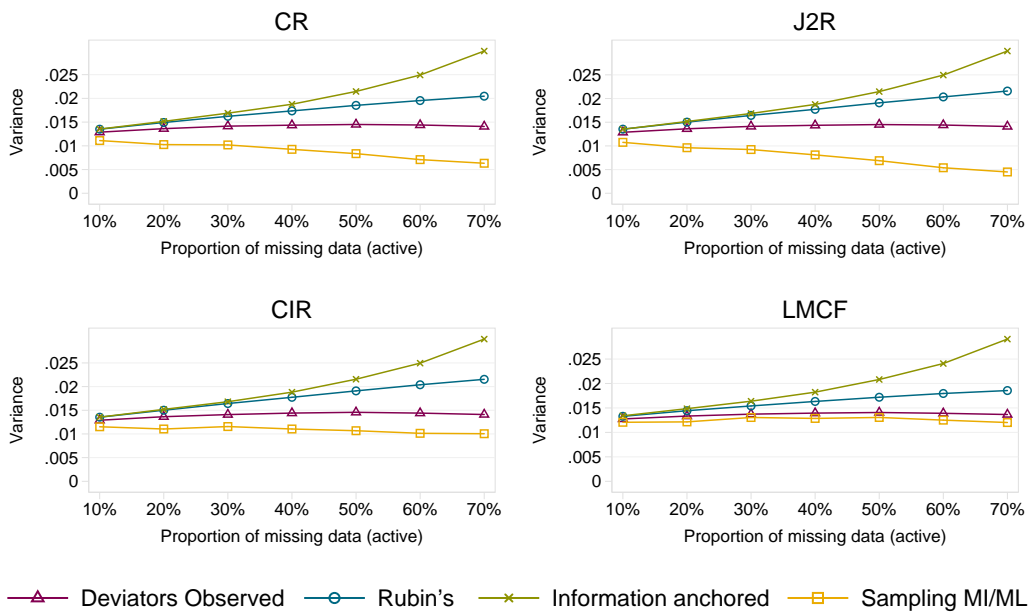


Figure C.4: Rubin's variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with $\Delta = 1$ and $n=100$ per arm.

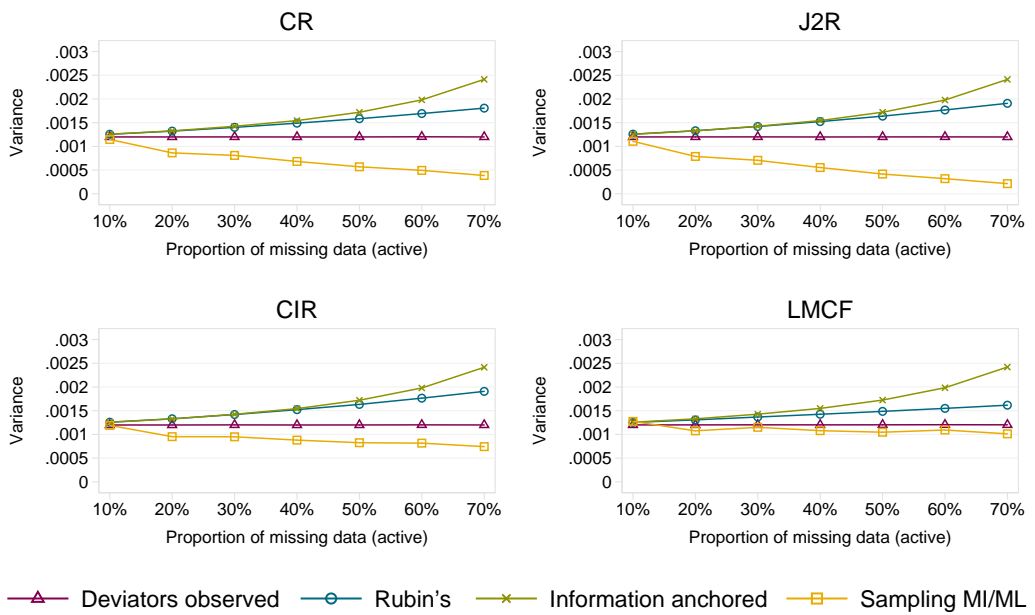


Figure C.5: Rubin's variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with $\Delta = 0$ and $n=1000$ per arm.

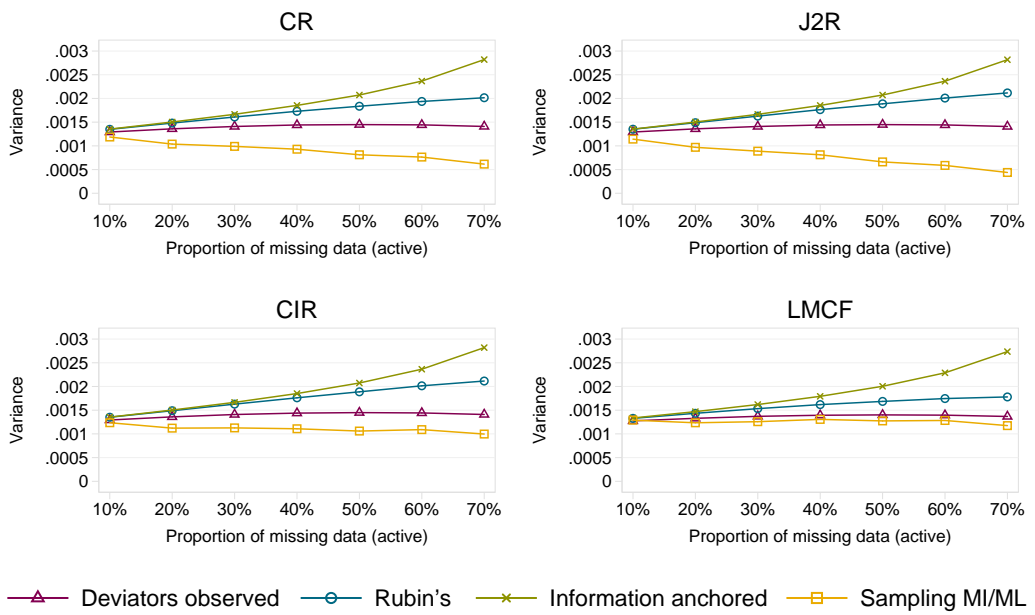


Figure C.6: Rubin's variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with $\Delta = 1$ and $n=1000$ per arm.

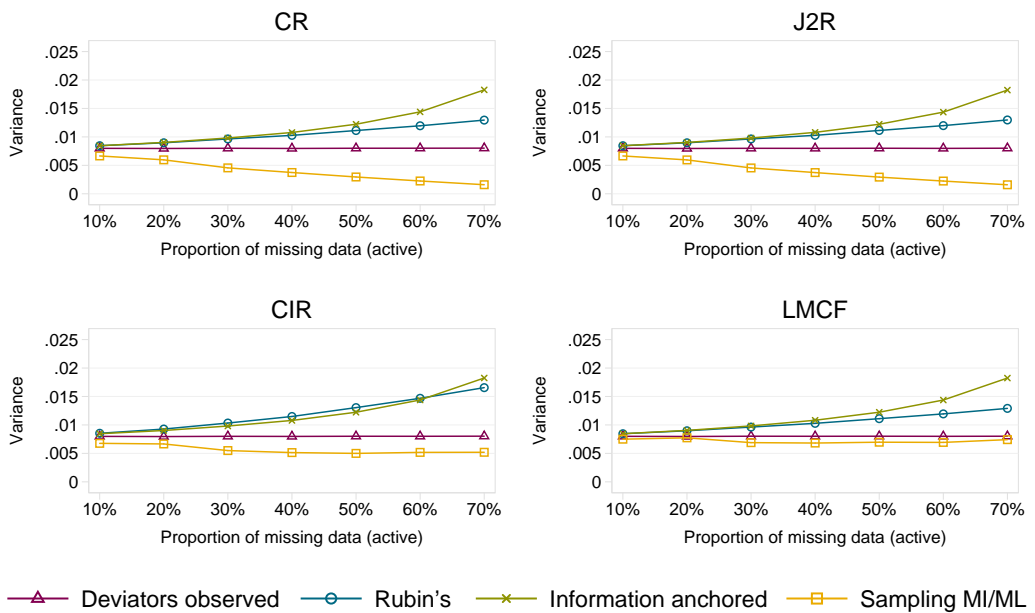


Figure C.7: Rubin's variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with $\Delta = 0$, $n=250$ per arm and a low covariance structure ($\sigma_{12} = 0.05$).

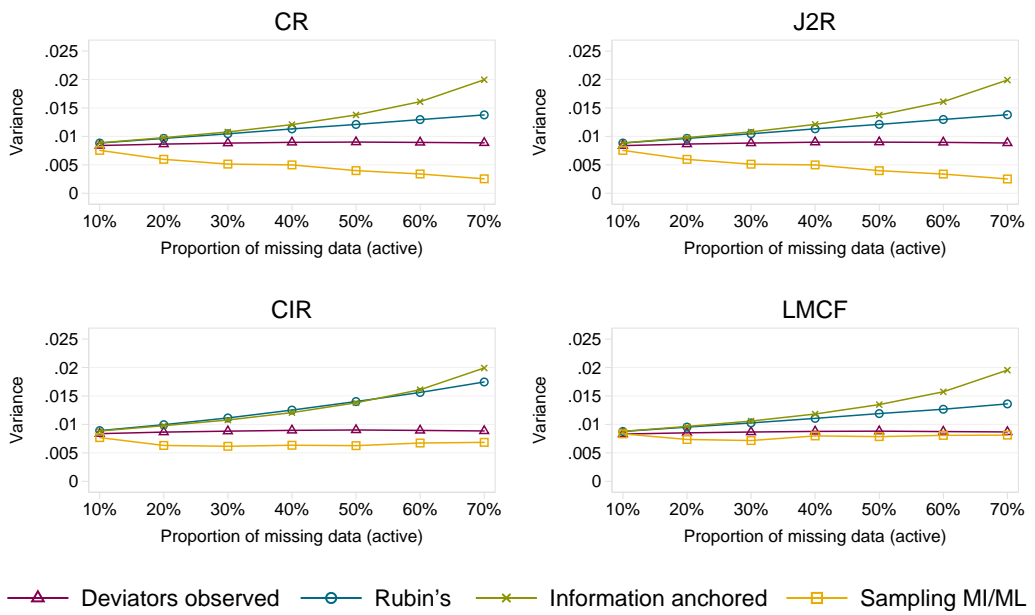


Figure C.8: Rubin's variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with $\Delta = 1$, $n=250$ per arm and a low covariance structure ($\sigma_{12} = 0.05$).

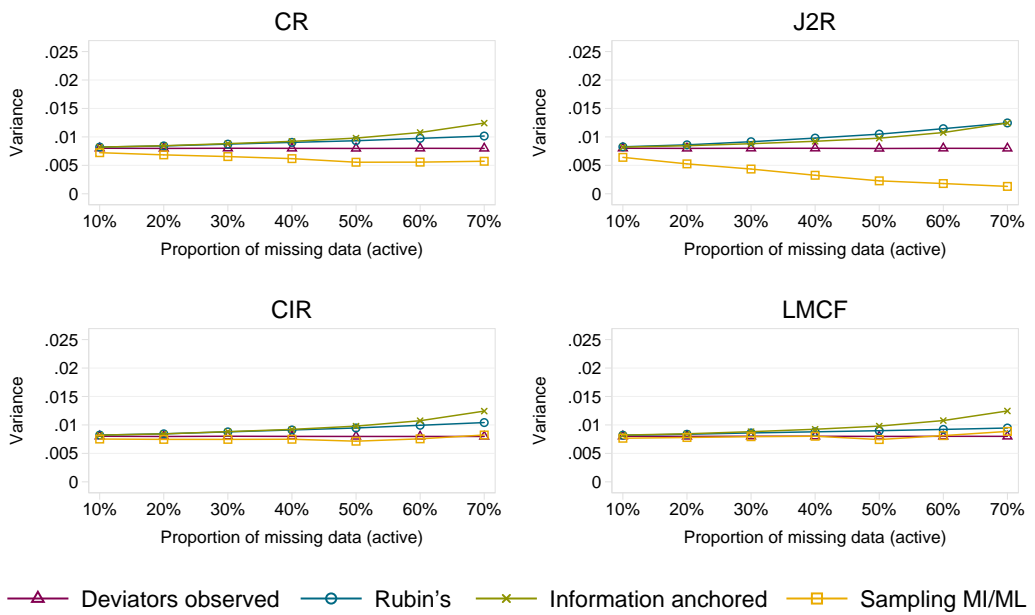


Figure C.9: Rubin's variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with $\Delta = 0$, $n=250$ per arm and a high covariance structure ($\sigma_{12} = 0.75$).

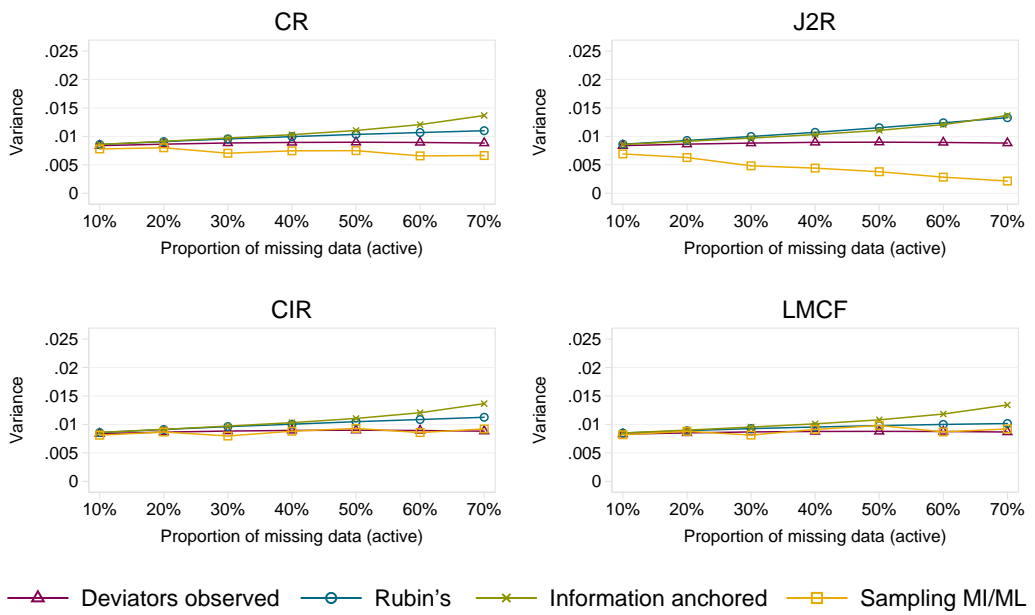


Figure C.10: Rubin's variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with $\Delta = 1$, $n=250$ per arm and a high covariance structure ($\sigma_{12} = 0.75$).

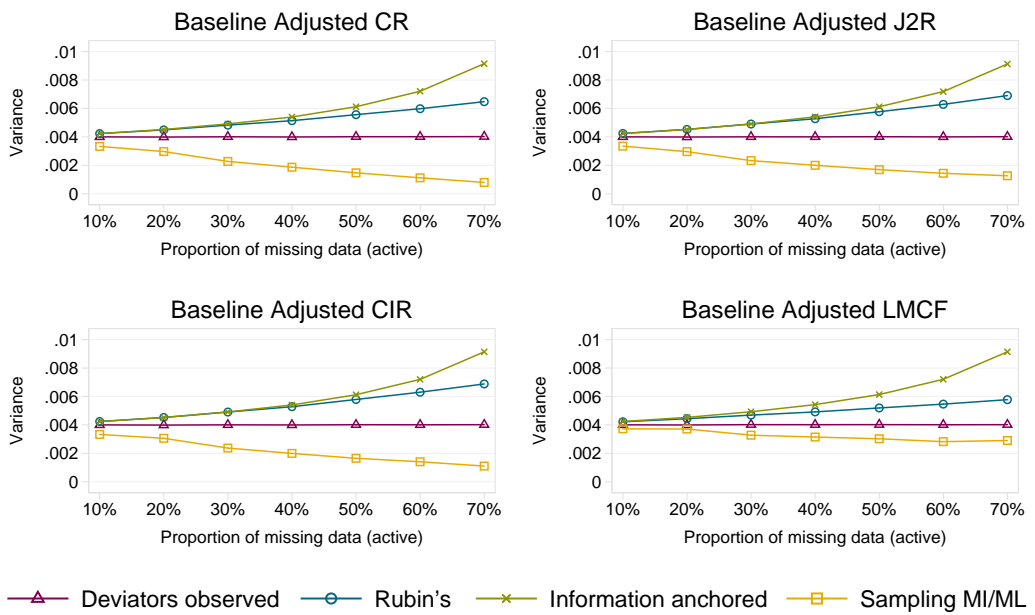


Figure C.11: Rubin's baseline adjusted variance estimator vs. the baseline adjusted information anchored variance vs. baseline adjusted variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 baseline adjusted MI estimates, with $\Delta = 0$ and $n=250$ per arm.

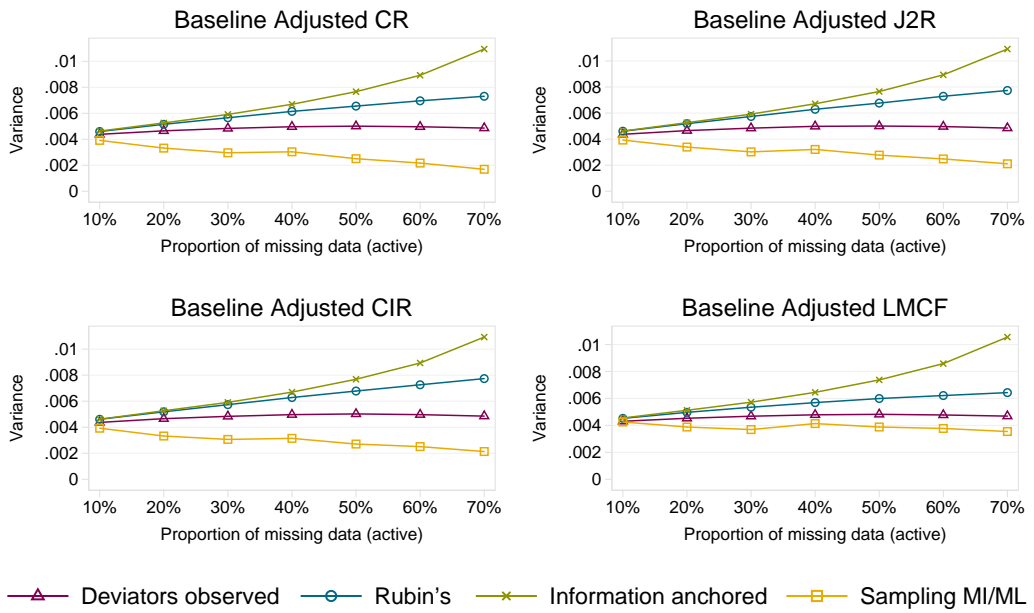


Figure C.12: Rubin's baseline adjusted variance estimator vs. the baseline adjusted information anchored variance vs. baseline adjusted variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 baseline adjusted MI estimates, with $\Delta = 1$ and $n=250$ per arm.

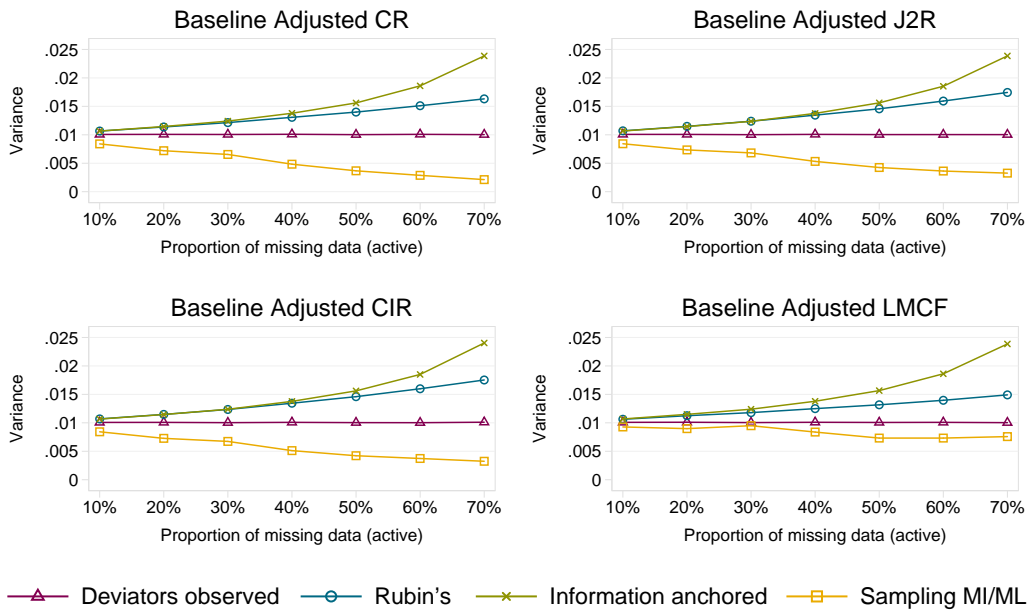


Figure C.13: Rubin's baseline adjusted variance estimator vs. the baseline adjusted information anchored variance vs. baseline adjusted variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 baseline adjusted MI estimates, with $\Delta = 0$ and $n=100$ per arm.

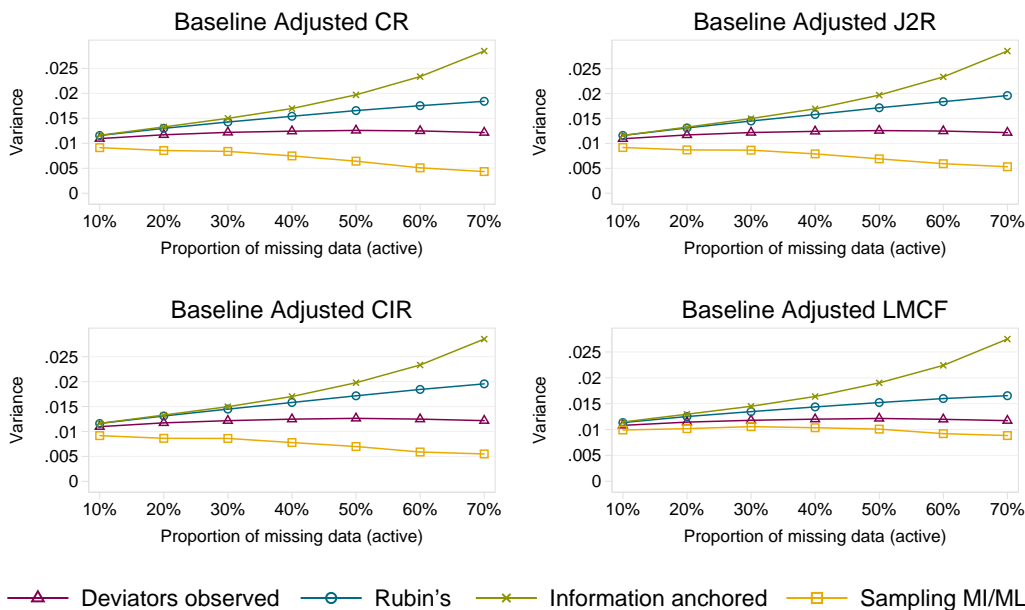


Figure C.14: Rubin's baseline adjusted variance estimator vs. the baseline adjusted information anchored variance vs. baseline adjusted variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 baseline adjusted MI estimates, with $\Delta = 1$ and $n=100$ per arm.

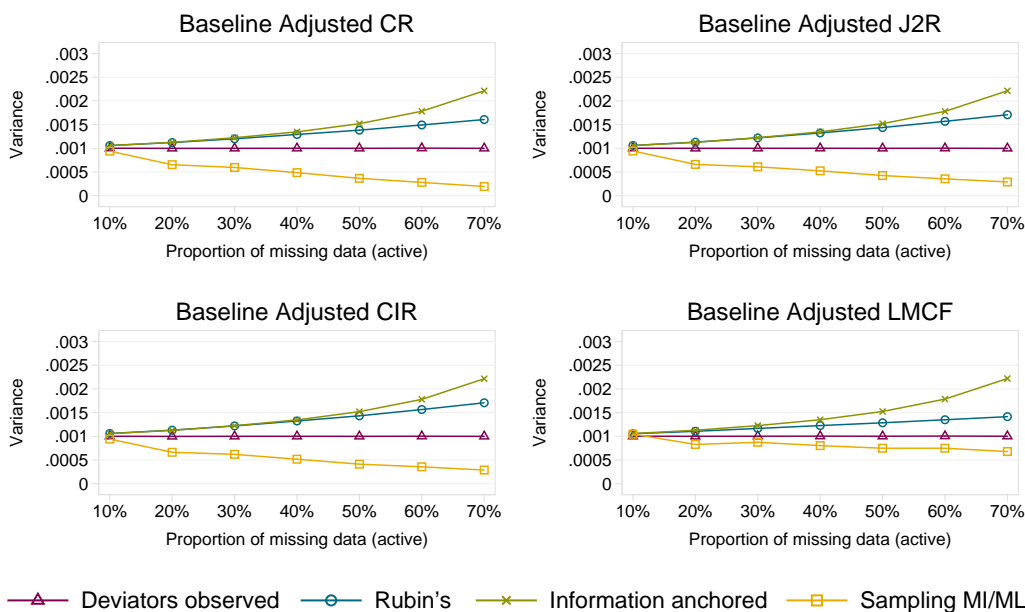


Figure C.15: Rubin's baseline adjusted variance estimator vs. the baseline adjusted information anchored variance vs. baseline adjusted variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 baseline adjusted MI estimates, with $\Delta = 0$ and $n=1000$ per arm.

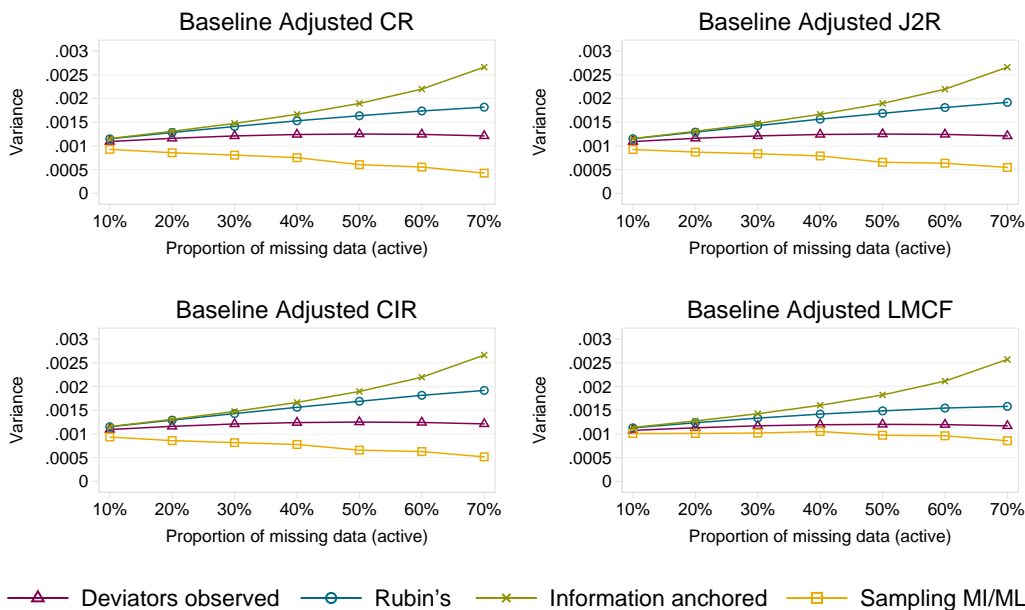


Figure C.16: Rubin's baseline adjusted variance estimator vs. the baseline adjusted information anchored variance vs. baseline adjusted variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 baseline adjusted MI estimates, with $\Delta = 1$ and $n=1000$ per arm.

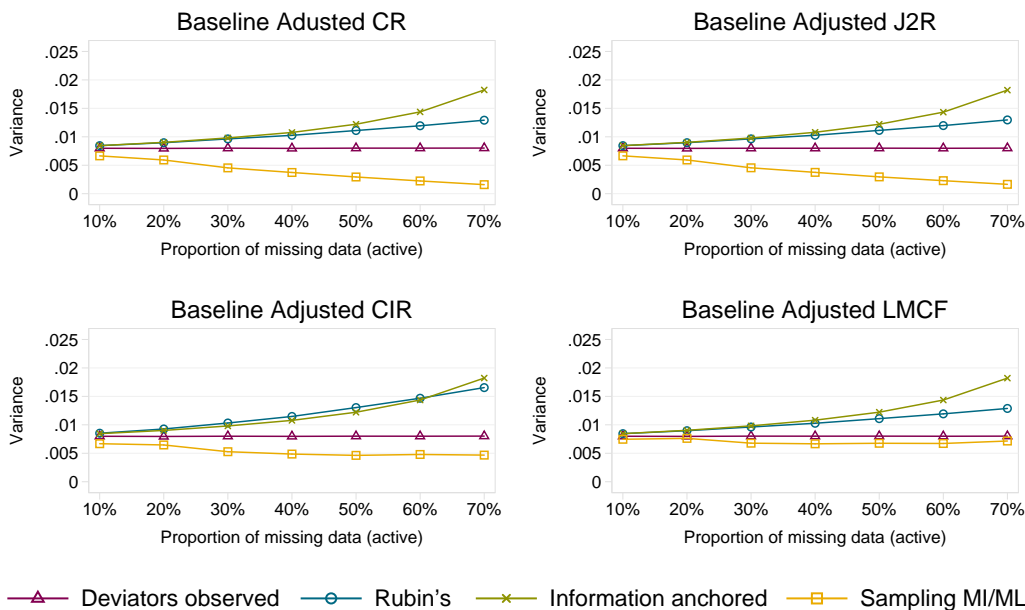


Figure C.17: Rubin's baseline adjusted variance estimator vs. the baseline adjusted information anchored variance vs. baseline adjusted variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 baseline adjusted MI estimates, with $\Delta = 0$, $n=250$ per arm and a low covariance structure ($\sigma_{12} = 0.05$).

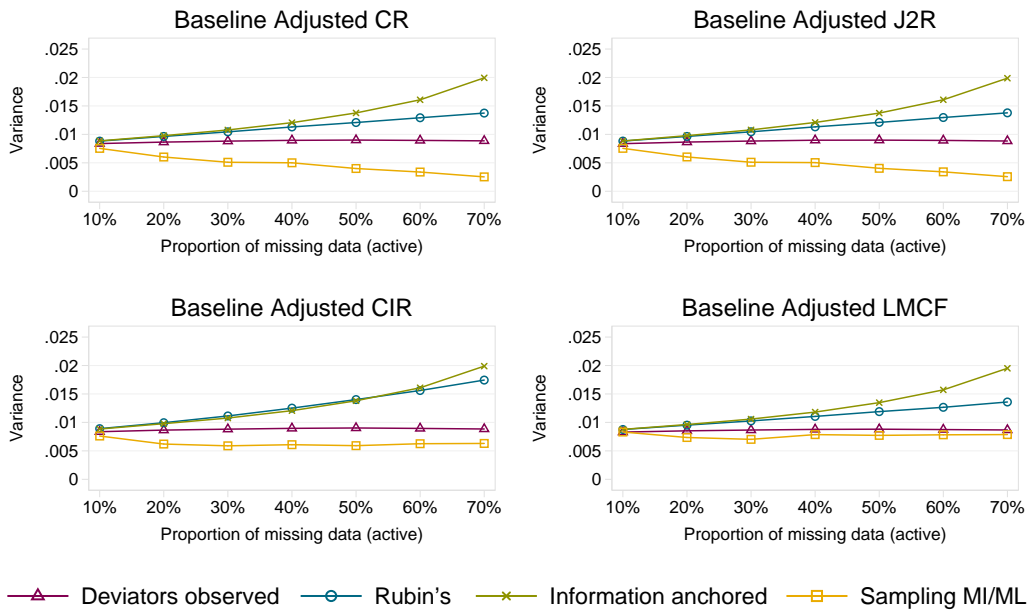


Figure C.18: Rubin's baseline adjusted variance estimator vs. the baseline adjusted information anchored variance vs. baseline adjusted variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 baseline adjusted MI estimates, with $\Delta = 1$, $n=250$ per arm and a low covariance structure ($\sigma_{12} = 0.05$).

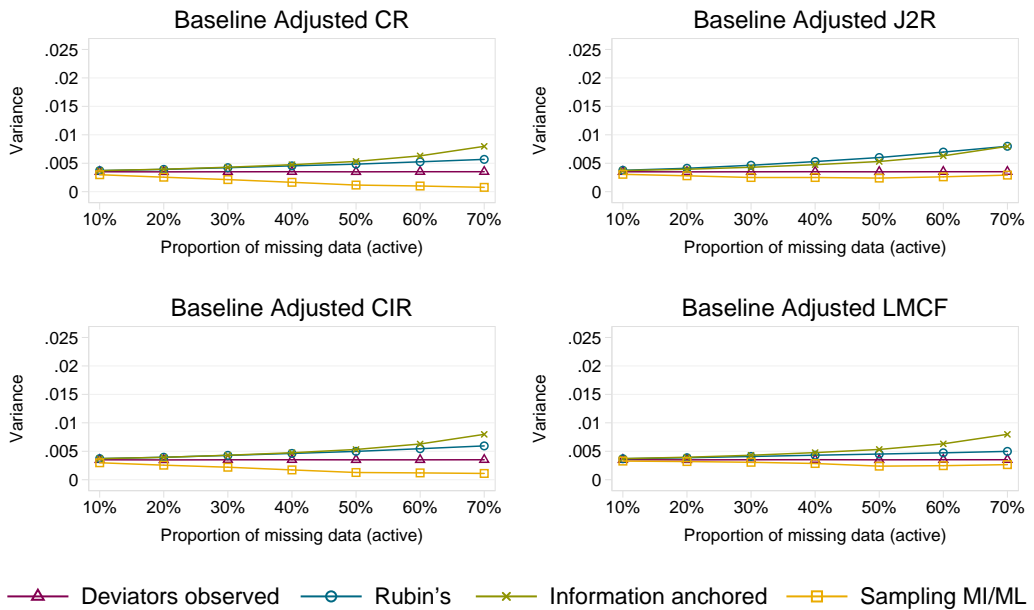


Figure C.19: Rubin's baseline adjusted variance estimator vs. the baseline adjusted information anchored variance vs. baseline adjusted variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 baseline adjusted MI estimates, with $\Delta = 0$, $n=250$ per arm and a high covariance structure ($\sigma_{12} = 0.75$).

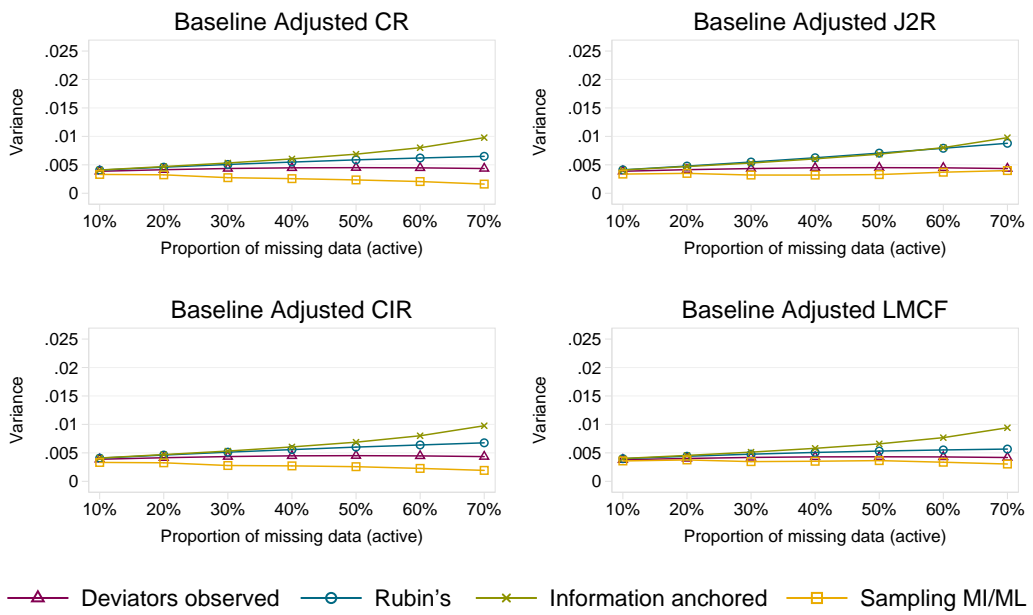


Figure C.20: Rubin's baseline adjusted variance estimator vs. the baseline adjusted information anchored variance vs. baseline adjusted variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 baseline adjusted MI estimates, with $\Delta = 1$, $n=250$ per arm and a high covariance structure ($\sigma_{12} = 0.05$).

Appendix D

Reference based computations; longitudinal setting

D.1 The design based variance estimator when post-deviation data is observed

D.1.1 Last measured variable subject to non-response

Consider a two arm trial comprising of n patients who have been randomised to an active arm ($z = a$) and n patients to a reference arm ($z = r$) with a continuous outcome measured repeatedly over a series of visits, $j = 1, \dots, J$ (including baseline) which we denote by Y_{zij} for $i = 1, \dots, n$. The mean treatment group difference at the final visit J is the primary outcome of interest. Suppose all n reference patients are observed at baseline and up to time J without deviation. However amongst the n active patients, only n_o patients are observed without deviation at all time points. The remaining n_d active arm patients deviate following time $J - 1$. For simplicity we assume there is no interim missing data. Let \mathcal{D} and \mathcal{O} define the sets of indices for the patients who do and do not deviate in the active arm respectively.

We assume the data comes from the following multivariate normal distributions with equal covariance structures (by arm),

$$\begin{pmatrix} Y_{zi1} \\ Y_{zi2} \\ \vdots \\ Y_{ziJ} \end{pmatrix} \sim N \left\{ \begin{pmatrix} \mu_{z1} \\ \mu_{z2} \\ \vdots \\ \mu_{zJ} \end{pmatrix}; \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1J} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1J} & \sigma_{2J} & \cdots & \sigma_{JJ} \end{bmatrix} \right\}, \quad i = 1, \dots, n \text{ for } z \in r \text{ or } i \in \mathcal{O} \text{ for } z \in a,$$

$$\begin{pmatrix} Y_{zi1} \\ Y_{zi2} \\ \vdots \\ Y_{ziJ} \end{pmatrix} \sim N \left\{ \begin{pmatrix} \mu_{z1} \\ \mu_{z2} \\ \vdots \\ \mu_{zJ} \end{pmatrix}; \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1J} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1J} & \sigma_{2J} & \cdots & \sigma_{JJ} \end{bmatrix} \right\}, i \in \mathcal{D}.$$

Interest lies in the mean treatment group difference at time J . The expectation of the treatment estimate at time J when the post-deviation data is observed is,

$$\left(\frac{n_o}{n} \mu_{aJ} + \frac{n_d}{n} \mu_{d,J} \right) - \mu_{r1}.$$

To establish the expectation of the design based variance estimate when post-deviation data is observed ($E[V_{\text{DF, full}}]$) we require the expectation of,

$$\frac{\frac{1}{n-1} \sum_{i=1}^n (Y_{riJ} - \bar{Y}_{rJ})^2}{n} + \frac{\frac{1}{n-1} \sum_{i=1}^n (Y_{aiJ} - \frac{n_o}{n} \bar{Y}_{aJ,o} - \frac{n_d}{n} \bar{Y}_{aJ,d})^2}{n},$$

where $\bar{Y}_{rJ} = \frac{1}{n} \sum_{i=1}^n Y_{riJ}$, $\bar{Y}_{aJ,o} = \frac{1}{n_o} \sum_{i \in \mathcal{O}} Y_{aiJ}$ and $\bar{Y}_{aJ,d} = \frac{1}{n_d} \sum_{i \in \mathcal{D}} Y_{aiJ}$. For the first term,

$$E \left[\frac{1}{n-1} \sum_{i=1}^n (Y_{riJ} - \bar{Y}_{rJ})^2 \right] = \sigma_{JJ}.$$

Following the same steps taken in Appendix B.1, for the second term we obtain,

$$E \left[\frac{1}{n-1} \sum_{i=1}^n \left(Y_{aiJ} - \frac{n_o}{n} \bar{Y}_{aJ,o} - \frac{n_d}{n} \bar{Y}_{aJ,d} \right)^2 \right] = \frac{1}{n-1} \left[(n-1) \left(\frac{n_o}{n} \sigma_{JJ} + \frac{n_d}{n} \sigma_{JJ} \right) + \frac{n_o n_d}{n} \Delta^2 \right],$$

where $\Delta = \mu_{aJ} - \mu_{dJ}$. Combining this result with the expectation of the first term assuming that n is sufficiently large so that we may take $(n-1)$ to be n gives,

$$E[V_{\text{DF, full}}] = \frac{2\sigma_{JJ}}{n} + \frac{n_o n_d}{n^3} \Delta^2.$$

In the de-jure setting where the deviators remain on-treatment $\mu_{dJ} = \mu_{aJ}$ and so we have,

$$E[V_{\text{DJ, full}}] = \frac{2\sigma_{JJ}}{n}.$$

For a typical RCT powered at 80% with 5% statistical significance $\Delta^2 \approx 15.68\sigma_{22}/n$ using (3.17). Thus as required, $E[V_{\text{DF, full}}] = E[V_{\text{DJ, full}}] + O(n^{-2})$.

Following the derivations in Section 3.4 we infer that in the current setting for the baseline adjusted treatment difference at time J ,

$$E[V_{\text{DF, full}}] = \frac{2\sigma_{J,1}}{n} + \frac{n_o n_d}{n^3} \Delta^2 \text{ and } E[V_{\text{DJ, full}}] = \frac{2\sigma_{J,1}}{n},$$

where $\sigma_{J,1} = \sigma_{JJ} - \frac{\sigma_{1J}^2}{\sigma_{11}}$. Thus in the baseline adjusted setting $E[V_{\text{DF, full}}] = E[V_{\text{DJ, full}}] + O(n^{-2})$ as required.

D.1.2 Monotone non-response

Suppose missingness is no longer restricted to the last follow-up time. Among the n_d deviators in the active arm we observe $n_{d,j}$ patients who deviate following time j , for $j = 1, \dots, J-1$ such that $n_d = \sum_{j=1}^{J-1} n_{d,j}$. Our assumptions for the observed data remain as in Appendix D.1.1. For the patients that deviate following time j we assume,

$$\begin{pmatrix} Y_{zi1} \\ Y_{zi2} \\ \vdots \\ Y_{ziJ} \end{pmatrix} \sim N \left\{ \begin{pmatrix} \mu_{z1} \\ \mu_{d,j,2} \\ \vdots \\ \mu_{d,j,J} \end{pmatrix}; \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1J} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1J} & \sigma_{2J} & \cdots & \sigma_{JJ} \end{bmatrix} \right\}, i \in \mathcal{DJ}.$$

Interest lies in the mean treatment group difference at time J . The expected value of the treatment estimate at time J when the post-deviation data is observed is,

$$\left(\frac{n_o}{n} \mu_{aJ} + \sum_{j=1}^{J-1} \frac{n_{d,j}}{n} \mu_{d,j,J} \right) - \mu_{r1}.$$

The variance of this estimate is calculated using,

$$\frac{\frac{1}{n-1} \sum_{i=1}^n (Y_{riJ} - \bar{Y}_{rJ})^2}{n} + \frac{\frac{1}{n-1} \sum_{i=1}^n \left(Y_{aiJ} - \frac{n_o}{n} \bar{Y}_{aJ,o} - \sum_{j=1}^{J-1} \frac{n_{d,j}}{n} \bar{Y}_{aJ,d,j} \right)^2}{n},$$

where $\bar{Y}_{rJ} = \frac{1}{n} \sum_{i=1}^n Y_{riJ}$, $\bar{Y}_{aJ,o} = \frac{1}{n_o} \sum_{i \in \mathcal{O}} Y_{aiJ}$ and $\bar{Y}_{aJ,d,j} = \frac{1}{n_{d,j}} \sum_{i \in \mathcal{DJ}} Y_{aiJ}$ for $j = 1, \dots, J-1$. Which, when assuming that n is sufficiently large so that we may take $(n-1)$ to be n has expected value,

$$E[V_{\text{DF, full}}] = \frac{\sigma_{JJ}}{n} + \frac{\sigma_{JJ}}{n} + \sum_{j=1}^{J-1} \frac{n_o n_{d,j} \Delta_{d,j}^2}{n^3} + \sum_{p=1}^{J-1} \sum_{q=1}^{(J-1, q \neq p)} \frac{n_{d,p} n_{d,q} (\Delta_{d,p,q})^2}{n^3},$$

where $\Delta_{d,j} = \mu_{aJ} - \mu_{d,j,J}$, $\Delta_{d,p,q} = \mu_{d,p,J} - \mu_{d,q,J}$, $\mu_{d,j,J}$ is the mean proposed under the de-facto scenario at time J , for patients who deviate following time j and $\mu_{d,p,J}$ and $\mu_{d,q,J}$ are the means

proposed under the de-facto scenario at time J , for patients who deviate following times p and q for $p = 1, \dots, J - 1$ and $q = 1, \dots, J - 1$.

In the de-jure setting where the deviators remain on-treatment $\mu_{d,j,J} = \mu_{a,J}$ for $j = 1, \dots, J$ and so we have,

$$E[V_{\text{DJ, full}}] = \frac{2\sigma_{JJ}}{n}.$$

For a typical RCT powered at 80% with 5% statistical significance $\Delta^2 \approx 15.68\sigma_{22}/n$ using (3.17). Thus as required, $E[V_{\text{DF, full}}] = E[V_{\text{DJ, full}}] + O(n^{-2})$.

Following the derivations in Section 3.4 we infer that in the current setting for the baseline adjusted treatment difference at time J ,

$$E[V_{\text{DF, full}}] = \frac{2\sigma_{J,1}}{n} + \sum_{j=1}^{J-1} \frac{n_{o}n_{d,j}\Delta_{d,j}^2}{n^3} + \sum_{p=1}^{J-1} \sum_{q=1}^{(J-1,q \neq p)} \frac{n_{d,p}n_{d,q}(\Delta_{d,p,q})^2}{n^3}$$

and $E[V_{\text{DJ, full}}] = \frac{2\sigma_{J,1}}{n}$,

where $\sigma_{J,1} = \sigma_{JJ} - \frac{\sigma_{1J}^2}{\sigma_{11}}$. Thus in the baseline adjusted setting $E[V_{\text{DF, full}}] = E[V_{\text{DJ, full}}] + O(n^{-2})$ as required.

Appendix E

Longitudinal simulation studies

E.1 Imposed missing data patterns

Scenario	Proportion of missing data		
	time 1	time 2	time 3
1	0%	5%	10%
2	0%	10%	20%
3	0%	15%	30%
4	0%	20%	40%
5	0%	25%	50%
6	0%	30%	60%
7	0%	35%	70%

Table E.1: Imposed missing data patterns for deviators in longitudinal simulations

Scenario	Proportion of missing data					
	time 1		time 2		time 3	
	active	reference	active	reference	active	reference
1	0%	0%	5%	2.5%	10%	5%
2	0%	0%	10%	5.0%	20%	10%
3	0%	0%	15%	7.5%	30%	15%
4	0%	0%	20%	10.0%	40%	20%
5	0%	0%	25%	12.5%	50%	25%
6	0%	0%	30%	15.0%	60%	30%
7	0%	0%	35%	17.5%	70%	35%

Table E.2: Imposed missing data patterns where deviation in reference arm was half that observed for the active arm in longitudinal simulations

E.2 Additional results

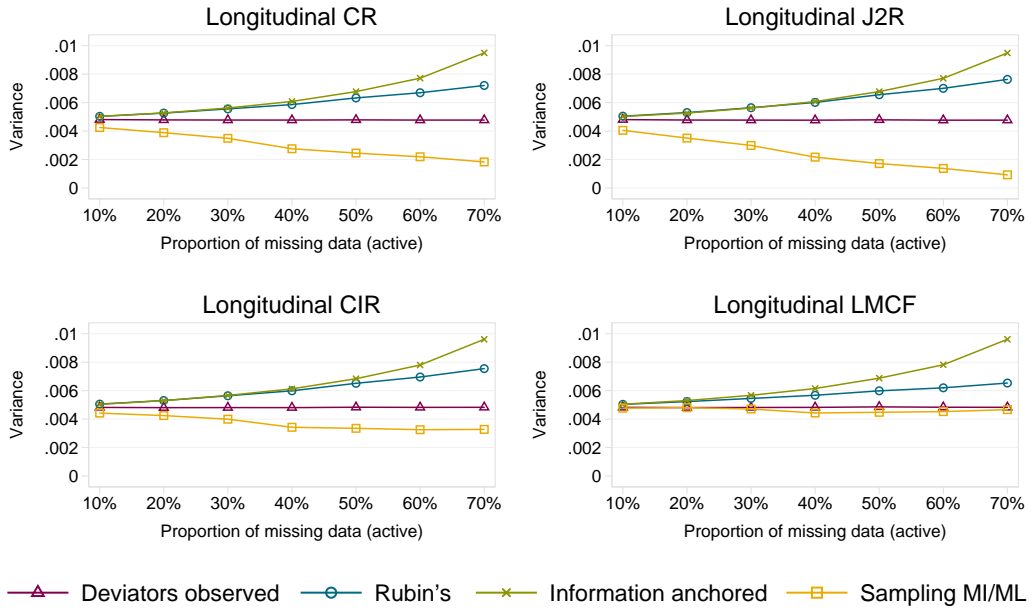


Figure E.1: Rubin's variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with $\Delta = 0$ and $n=250$ per arm.

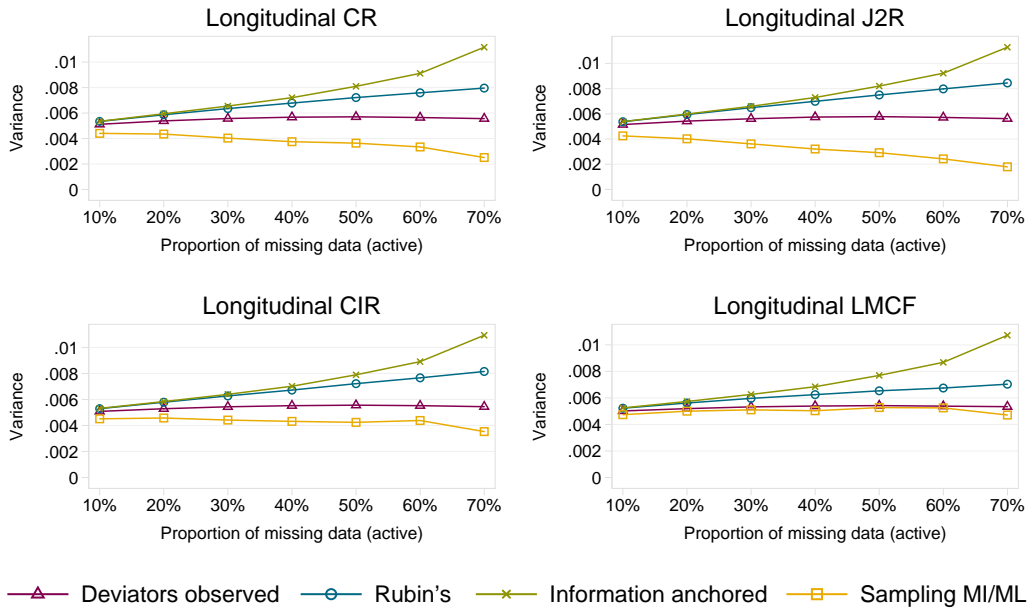


Figure E.2: Rubin's variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with $\Delta = 1$ and $n=250$ per arm.

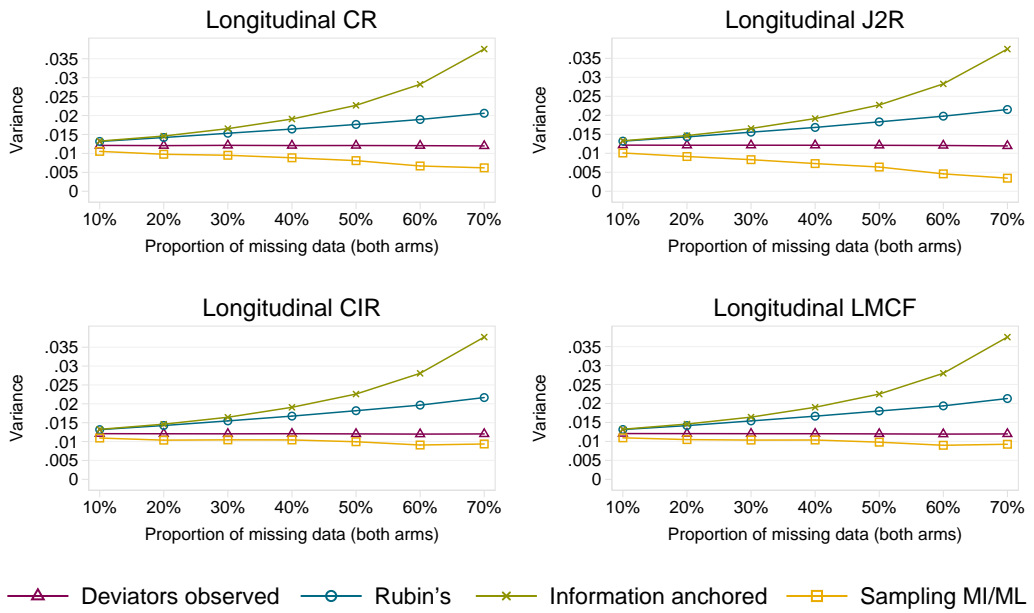


Figure E.3: Rubin's variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with the asthma trial treatment effect, $\Delta = 0.3$, deviation in both treatment arms and $n=100$ per arm.

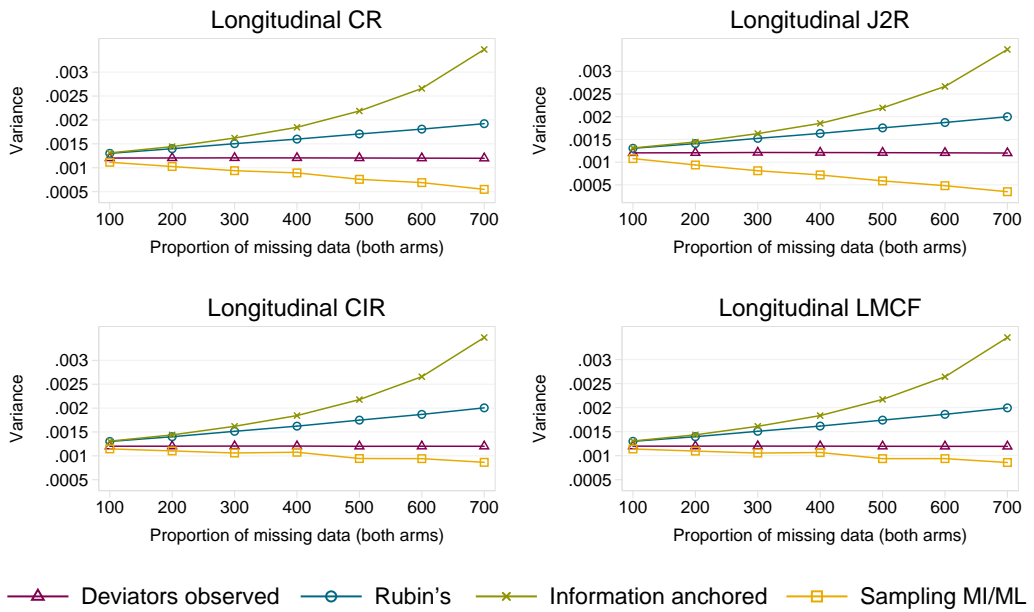


Figure E.4: Rubin's variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with the asthma trial treatment effect, $\Delta = 0.3$, deviation in both treatment arms and $n=1000$ per arm.

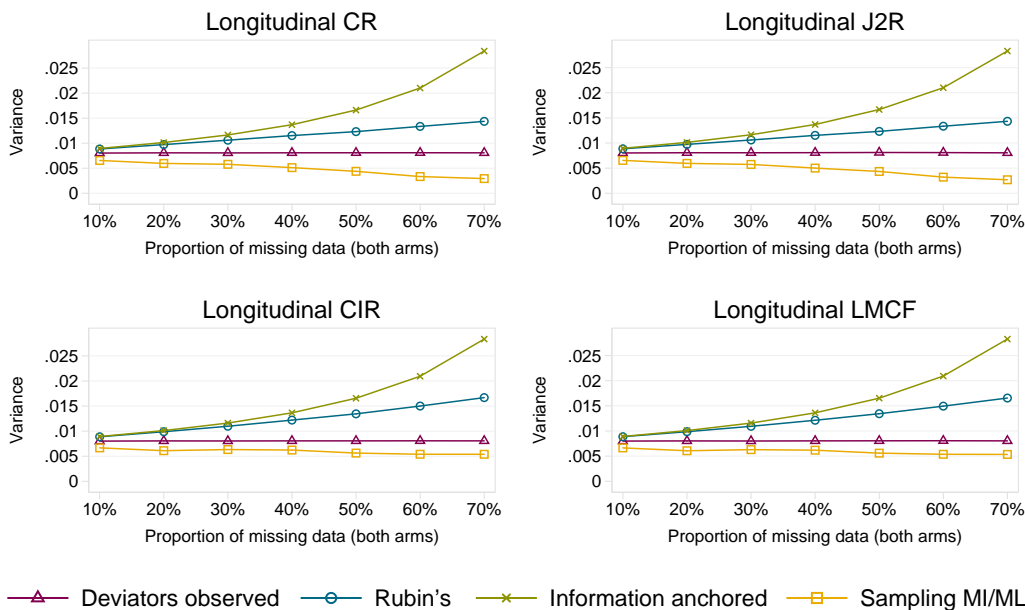


Figure E.5: Rubin's variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with the asthma trial treatment effect, $\Delta = 0.3$, deviation in both treatment arms, $n=250$ per arm and a low covariance structure ($\sigma_{j,j^*} = 0.05$).

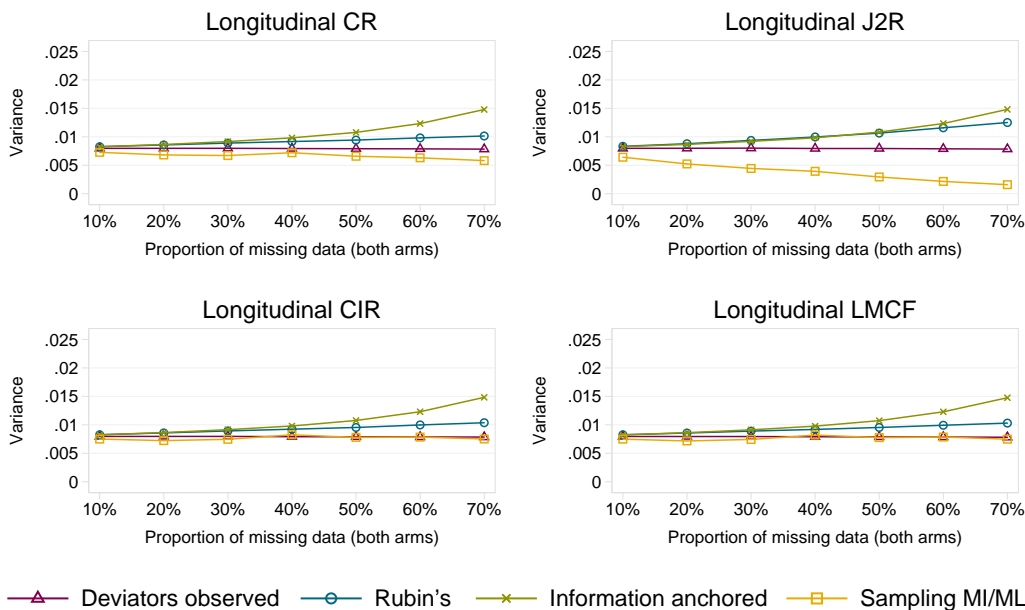


Figure E.6: Rubin's variance estimator vs. information anchored variance vs. variance where deviators observed (all averaged over 1000 simulations) vs. long-run sampling variance of the 1000 MI estimates, with the asthma trial treatment effect, $\Delta = 0.3$ deviation in both treatment arms, $n=250$ per arm and a high covariance structure ($\sigma_{j,j^*} = 0.75$).

Appendix F

The ‘ δ -method’ computations

F.1 The design based variance estimator when post-deviation data is observed; baseline and single follow-up

Generally, when post-deviation data is observed on-treatment $\mathbf{a}^T \mathbf{D}_{dj} \mathbf{Y}$ returns the de-jure treatment estimate and

$$E [V (\mathbf{a}^T \mathbf{D}_{dj} \mathbf{Y})] = E [\mathbf{a}^T \mathbf{D}_{dj} V (\mathbf{Y}) \mathbf{D}_{dj}^T \mathbf{a}] = \mathbf{a}^T \mathbf{D}_{dj} \Sigma \mathbf{D}_{dj}^T \mathbf{a}.$$

When post-deviation data is observed with a δ -adjustment, $\mathbf{a}^T \mathbf{D}_\delta \mathbf{Y}$ returns the δ -adjusted treatment estimate and

$$E [V (\mathbf{a}^T \mathbf{D}_\delta \mathbf{Y})] = E [\mathbf{a}^T \mathbf{D}_\delta V (\mathbf{Y}) \mathbf{D}_\delta^T \mathbf{a}] = \mathbf{a}^T \mathbf{D}_\delta \Sigma \mathbf{D}_\delta^T \mathbf{a}.$$

In the two arm trial setting ($z = a, r$) with baseline (time 1) and a single follow-up outcome (time 2) interest lies in the unadjusted mean treatment group difference at time 2 where Y_{zi2} denote the continuous outcome measure for patients at time 2 in arm z for $i = 1, \dots, n$. We assume $Y_{zi2} \sim N(\mu_z, \sigma_{22})$ however n_d active arm patients deviate such that $Y_{ai2} \sim N(\mu_d, \sigma_{22})$ for $i \in \mathcal{D}$.

Following the steps in Appendix B.1, for when it is proposed that $\mu_{d2} = \mu_a + \delta$ the expected variance of the treatment estimate is,

$$E [V_{\delta, \text{full}}] = \mathbf{a}^T \mathbf{D}_\delta \Sigma \mathbf{D}_\delta^T \mathbf{a} = \frac{2\sigma_{22}}{n} + \frac{n_o n_d \delta^2}{n^3}.$$

Under de-jure/on-treatment behaviour for deviators, $\mu_{d2} = \mu_a$ thus,

$$E[V_{\text{DJ, full}}] = \mathbf{a}^T \mathbf{D}_{\text{dj}} \Sigma \mathbf{D}_{\text{dj}}^T \mathbf{a} = \frac{2\sigma_{22}}{n},$$

and we see $V_{\delta, \text{full}} = V_{\text{DJ, full}} + O(\pi_d(1 - \pi_d)/n)$.

Following the derivations in Section 3.4 we infer that in the current setting for the baseline adjusted treatment difference at time 2,

$$E[V_{\delta, \text{full}}] = \frac{2\sigma_{2.1}}{n} + \frac{n_o n_d}{n^3} \delta^2 \text{ and } E[V_{\text{DJ, full}}] = \frac{2\sigma_{2.1}}{n},$$

where $\sigma_{2.1} = \sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}$. Thus in the baseline adjusted setting $E[V_{\delta, \text{full}}] = E[V_{\text{DJ, full}}] + O(\pi_d(1 - \pi_d)/n)$ as required.

F.2 The design based variance estimator when post-deviation data is unobserved; imputation calculations

Here we present additional details of the calculations involved in establishing the expectation of Rubin's MI variance estimator in the δ -adjusted case in Section 5.3. As presented in the main text we require $E(V_{\delta, \text{MI}}) = E(\hat{W}) + (1 + \frac{1}{K}) E(\hat{B})$. First we consider $E(\hat{W})$,

$$E(\hat{W}) = E(\hat{\sigma}_k^2) = \frac{E(\hat{\sigma}_{22,r})}{n} + \frac{E(\hat{\sigma}_{22,a,k})}{n},$$

where,

$$(n-1) E(\hat{\sigma}_{22,r}) = E\left(\sum_{i=1}^n (Y_{ri2} - \hat{\mu}_{r2})^2\right) = (n-1)\sigma_{22},$$

$$(n-1) E(\hat{\sigma}_{22,a,k}) = E\left[\sum_{i \in \mathcal{O}} (Y_{ai2} - \hat{\mu}_{a2,k})^2\right] + E\left[\sum_{i \in \mathcal{D}} (\hat{Y}_{ai2,k} - \hat{\mu}_{a2,k})^2\right].$$

Decomposing the two-right hand components separately, over the observed cases we have,

$$\begin{aligned} & E\left[\sum_{i \in \mathcal{O}} (Y_{ai2} - \hat{\mu}_{a2,k})^2\right] \\ &= E\left[\sum_{i \in \mathcal{O}} \left(Y_{ai2} - \frac{n_o}{n} \bar{Y}_{a2o} - \frac{n_d}{n} \bar{Y}_{a2d,k}\right)^2\right] \\ &= E\left[\sum_{i \in \mathcal{O}} \left(Y_{ai2} - \frac{n_o}{n} \bar{Y}_{a2o} - \frac{n_d}{n} \left[\bar{Y}_{a2o} + \delta_k + u_k + \left(\frac{r}{q} + b_k\right) (\bar{Y}_{a1d} - \bar{Y}_{a1o}) + \bar{e}_k\right]\right)^2\right] \\ &= E\left[\sum_{i \in \mathcal{O}} \left((Y_{ai2} - \bar{Y}_{a2o}) - \frac{n_d}{n} \delta_k - \frac{n_d}{n} u_k - \frac{n_d}{n} \left(\frac{r}{q} + b_k\right) (\bar{Y}_{a1d} - \bar{Y}_{a1o}) - \frac{n_d}{n} \bar{e}_k\right)^2\right]. \end{aligned}$$

Expanding the square and taking expectation term by term results in the same non-zero components as in the MAR setting outlined in Section B.4.1, along with the following component,

$$E \left[\sum_{i \in \mathcal{O}} \left(\frac{n_d}{n} \right)^2 \delta_k^2 \right] = \left(\frac{n_d}{n} \right)^2 n_o (\delta + \sigma_\delta^2).$$

Over the deviators we have,

$$\begin{aligned} & E \left[\sum_{i \in \mathcal{D}} \left(\hat{Y}_{ai2,k} - \hat{\mu}_{a2,k} \right)^2 \right] \\ &= E \left[\sum_{i \in \mathcal{D}} \left(\bar{Y}_{a2o} + \delta_k + u_k + \left(\frac{r}{q} + b_k \right) (Y_{ai1} - \bar{Y}_{a1o}) + e_{i,k} - \frac{n_o}{n} \bar{Y}_{a2o} - \frac{n_d}{n} \bar{Y}_{a2d,k} \right)^2 \right] \\ &= E \left[\sum_{i \in \mathcal{D}} \left(\frac{n_o}{n} \delta_k + \frac{n_o}{n} u_k + \left(\frac{r}{q} + b_k \right) (Y_{ai1} - \bar{Y}_{a1o}) - \frac{n_d}{n} \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{a1o}) - \frac{n_d}{n} \bar{e}_k + e_{i,k} \right)^2 \right]. \end{aligned}$$

Expanding the square and taking expectation term by term results in the same non-zero components as in the MAR setting outlined in Section B.4.1, along with the following component,

$$E \left[\sum_{i \in \mathcal{D}} \left(\frac{n_o}{n} \right)^2 \delta_k^2 \right] = \left(\frac{n_o}{n} \right)^2 n_d (\delta + \sigma_\delta^2).$$

Combining the results over the observed and deviating cases for time 1 a baseline where $\mu_{a1} = \mu_{r1}$ results in (5.2). For $(1 + \frac{1}{K}) E(\hat{B})$ we require,

$$E[\hat{B}] = \frac{1}{K-1} E \left[\sum_{k=1}^K \left(\hat{\theta}_{\delta,k} - \hat{\theta}_{\delta, \text{MI}} \right)^2 \right],$$

where,

$$\hat{\theta}_{\delta,k} = \frac{n_o}{n} \bar{Y}_{a2o} + \frac{n_d}{n} \left[\bar{Y}_{a2o} + \delta_k + u_k + \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{a1o}) + \bar{e}_k \right] - \bar{Y}_{r2},$$

$$\hat{\theta}_{\delta, \text{MI}} = \frac{1}{K} \sum_{k=1}^K \hat{\theta}_{\delta,k}.$$

That is,

$$\begin{aligned} E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \delta_k - \frac{n_d}{n} \bar{\delta} + \frac{n_d}{n} u_k - \frac{n_d}{n} \bar{u} + \frac{n_d}{n} \left(\frac{r}{q} + b_k \right) (\bar{Y}_{a1d} - \bar{Y}_{a1o}) - \frac{n_d}{n} \left(\frac{r}{q} + \bar{b} \right) (\bar{Y}_{a1d} - \bar{Y}_{a1o}) \right. \right. \\ \left. \left. + \frac{n_d}{n} \bar{e}_k - \frac{n_d}{n} \bar{e} \right)^2 \right]. \end{aligned}$$

Expanding the square and taking expectation by term gives the same non-zero components as in the MAR setting outlined in Section B.4.1, along with the following components,

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \delta_k^2 \right] = K \left(\frac{n_d}{n} \right)^2 (\delta^2 + \sigma_\delta^2),$$

$$E \left[\sum_{k=1}^K \left(\frac{n_d}{n} \right)^2 \bar{\delta}^2 \right] = \left(\frac{n_d}{n} \right)^2 K \left(\delta^2 + \frac{\sigma_\delta^2}{K} \right),$$

$$E \left[\sum_{k=1}^K -2 \left(\frac{n_d}{n} \right)^2 \delta_k \bar{\delta} \right] = -2 \left(\frac{n_d}{n} \right)^2 K \left(\delta^2 + \frac{\sigma_\delta^2}{K} \right).$$

Combining the results gives (5.3).

Proportion of Missing data	Derived Variance	Average Variance	MSCE
10% (N=25)	0.00506	0.00507	0.000011
20% (N=50)	0.00543	0.00544	0.000013
30% (N=75)	0.00593	0.00596	0.000015
40% (N=100)	0.00660	0.00664	0.000021
50% (N=125)	0.00752	0.00760	0.000029
60% (N=150)	0.00885	0.00892	0.000040
70% (N=175)	0.01092	0.01114	0.000062

Table F.1: Derived Rubin’s variance versus simulated Rubin’s variance (averaged over 1000 simulations) with $\delta_k \sim N(0.1, 0.05^2)$ adjustment in the asthma RCT inspired setting with $K=50$.

Proportion of Missing data	Derived Variance	Average Variance	MSCE
10% (N=25)	0.00503	0.00503	0.000011
20% (N=50)	0.00532	0.00532	0.000012
30% (N=75)	0.00569	0.00570	0.000014
40% (N=100)	0.00619	0.00624	0.000019
50% (N=125)	0.00688	0.00693	0.000024
60% (N=150)	0.00792	0.00804	0.000036
70% (N=175)	0.00967	0.00983	0.000054

Table F.2: Derived Rubin’s variance versus simulated Rubin’s variance (averaged over 1000 simulations) with fixed $\delta_k = 0.1$ adjustment in the asthma RCT inspired setting with $K=50$.

F.3 The design based variance estimator when post-deviation data is observed; longitudinal setting

Consider the baseline (time 1) and J follow-up setting with $n_{d,j}$ patients who deviate following time j , for $j = 1, \dots, J - 1$ such that $n_d = \sum_{j=1}^{J-1} n_{d,j}$. Interest lies in the unadjusted mean treatment group difference at time J . Following the steps outlined in Appendix D.1.2, the expected value of the treatment estimate at time J when post-deviation δ -adjusted data is observed is,

$$\left(\frac{n_o}{n} \mu_{aJ} + \sum_{j=1}^{J-1} \frac{n_{d,j}}{n} \mu_{d,j,J} \right) - \mu_{r1},$$

where $\mu_{d,j,J} = \mu_{aJ} + (J - j)\delta$. The variance of this estimate is calculated using,

$$\frac{\frac{1}{n-1} \sum_{i=1}^n (Y_{riJ} - \bar{Y}_{rJ})^2}{n} + \frac{\frac{1}{n-1} \sum_{i=1}^n \left(Y_{aiJ} - \frac{n_o}{n} \bar{Y}_{aJ,o} - \sum_{j=1}^{J-1} \frac{n_{d,j}}{n} \bar{Y}_{aJ,d,j} \right)^2}{n},$$

where $\bar{Y}_{rJ} = \frac{1}{n} \sum_{i=1}^n Y_{riJ}$, $\bar{Y}_{aJ,o} = \frac{1}{n_o} \sum_{i \in \mathcal{O}} Y_{aiJ}$ and $\bar{Y}_{aJ,d,j} = \frac{1}{n_{d,j}} \sum_{i \in \mathcal{D}j} Y_{aiJ}$ for $j = 1, \dots, J - 1$. Which, when assuming that n is sufficiently large so that we may take $(n - 1)$ to be n has expected value,

$$E[V_{\delta, \text{full}}] = \frac{\sigma_{JJ}}{n} + \frac{\sigma_{JJ}}{n} + \sum_{j=1}^{J-1} \frac{n_o n_{d,j} (J-j)^2 \delta^2}{n^3} + \sum_{p=1}^{J-1} \sum_{q=1}^{(J-1, q \neq p)} \frac{n_{d,p} n_{d,q} ((J-p)\delta - (J-q)\delta)^2}{n^3}.$$

In the de-jure setting where the deviators remain on-treatment $\delta = 0$ for $j = 1, \dots, J$ and so we have,

$$E[V_{D,J, \text{full}}] = \frac{2\sigma_{JJ}}{n}.$$

Thus as required, $E[V_{\delta, \text{full}}] = E[V_{D,J, \text{full}}] + O(\pi_{d,j}(1 - \pi_d)n^{-1}) + O(\pi_{d,p}\pi_{d,p}n^{-1})$. Following the derivations in Section 3.4 we infer that in the current setting for the baseline adjusted treatment difference at time J ,

$$E[V_{\delta, \text{full}}] = \frac{2\sigma_{J,1}}{n} + \sum_{j=1}^{J-1} \frac{n_o n_{d,j} (J-j)^2 \delta^2}{n^3} + \sum_{p=1}^{J-1} \sum_{q=1}^{(J-1, q \neq p)} \frac{n_{d,p} n_{d,q} ((J-p)\delta - (J-q)\delta)^2}{n^3}$$

and $E[V_{\delta, \text{full}}] = \frac{2\sigma_{J,1}}{n}$, where $\sigma_{J,1} = \sigma_{JJ} - \frac{\sigma_{JJ}^2}{\sigma_{11}}$. Thus in the baseline adjusted setting $E[V_{D,F, \text{full}}] = E[V_{D,J, \text{full}}] + O(\pi_{d,j}(1 - \pi_d)n^{-1}) + O(\pi_{d,p}\pi_{d,p}n^{-1})$ as required.

Bibliography

- [1] White, I. R., Horton, N. J., Carpenter, J. R., and Pocock, S. J., “Strategy for intention to treat analysis in randomised trials with missing outcome data,” *British Medical Journal*, vol. 342, 2011.
- [2] Satelli, A., Chan, K., and Scott, E. M., *Sensitivity analysis*. Chichester: Wiley, 2000.
- [3] Satelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S., *Global Sensitivity Analysis: The Primer*. New York: Wiley, 2009.
- [4] Daniels, M. J. and Hogan, J. W., *Missing Data in Longitudinal Studies, Strategies for Bayesian Modelling and Sensitivity Analysis*. Monographs on Statistics and Applied Probability, Chapman and Hall, 2008.
- [5] Bell, M., Fiero, M., Horton, N. J., and Hsu, C.-H., “Handling missing data in rcts, a review of the top medical journals,” *BMC Medical Research Methodology*, vol. 14, no. 1, p. 118, 2014.
- [6] Molenberghs, G. and Kenward, M. G., *Missing Data in Clinical Studies*. Wiley, 2007.
- [7] Carpenter, J. R. and Kenward, M. G., “Missing data in randomised controlled trials - a practical guide.,” Available from http://missingdata.lshtm.ac.uk/downloads/rm04_jh17_mk.pdf, accessed June 2014, 2008.
- [8] Molenberghs, G., Fitzmaurice, G., Kenward, M. G., Tsiatis, A., and Verbeke, G., *Handbook of Missing Data Methodology*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods, Taylor & Francis, 2014.
- [9] Wood, A. M., White, I. R., and Thompson, S. G., “Are missing outcome data adequately handled? a review of published randomized controlled trials in major medical journals,” *Clinical Trials*, vol. 1, no. 4, pp. 368–376, 2004.
- [10] Wood, A. M., White, I. R., Hillsdon, M., and Carpenter, J. R., “Comparison of imputation and modelling methods in the analysis of a physical activity trial with missing outcomes,” *International Journal of Epidemiology*, vol. 34, no. 1, pp. 89–99, 2005.
- [11] Committee for Medicinal Products for Human Use, *Guideline on missing data in confirmatory clinical trials*. London, UK: European Medicines Agency, 2010.

- [12] National Research Council, *The Prevention and Treatment of Missing Data in Clinical Trials*. Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioural and Social Sciences Education., Washington, DC: The National Academies Press, 2010.
- [13] Carpenter, J. R., Roger, J. H., and Kenward, M. G., “Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions and inference via multiple imputation,” *Journal of Biopharmaceutical Statistics*, vol. 23, no. 6, pp. 1352–1371, 2013.
- [14] Ma, Y., Olendzki, B. C., Wang, J., Pursutte, G. M., Li, W., Fang, H., and et al., “Single-component versus multicomponent dietary goals for the metabolic syndrome: a randomized trial,” *Ann Intern Med*, vol. 162, no. 4, pp. 248–257, 2015.
- [15] Grobler, A. C., Matthews, G., and Molenberghs, G., “The impact of missing data on clinical trials: A re-analysis of a placebo controlled trial of hypericum perforatum (st johns wort) and sertraline in major depressive disorder,” *Psychopharmacology*, vol. 231, no. 9, pp. 1987–1999, 2014.
- [16] Kenward, M. G., “Controlled multiple imputation methods for sensitivity analyses in longitudinal clinical trials with dropout and protocol deviation,” *Clinical Investigation*, vol. 5, no. 3, pp. 311–320, 2015.
- [17] Seaman, S. R., White, I. R., and Leacy, F. P., “Comment on ‘analysis of longitudinal trials with protocol deviations— a framework for relevant, accessible assumptions, and inference via multiple imputation’ by carpenter, roger and kenward,” *Journal of Biopharmaceutical Statistics*, vol. 24, no. 6, pp. 1358–1362, 2014.
- [18] Rubin, D. B., “Inference and missing data,” *Biometrika*, vol. 63, no. 3, pp. 581–592, 1976.
- [19] Rubin, D. B., *Multiple Imputation for Nonreponse in Surveys*. Wiley Series in Probability and Mathematical Statistics, USA: John Wiley & Sons, 1987.
- [20] Seaman, S. R. and White, I. R., “Review of inverse probability weighting for dealing with missing data,” *Statistical methods in medical research*, vol. 22, no. 3, pp. 278–295, 2011.
- [21] Carpenter, J. R., Kenward, M. G., and Vansteelandt, S., “A comparison of multiple imputation and doubly robust estimation for analyses with missing data,” *JRSS*, vol. 169, no. 3, pp. 571–584, 2006.
- [22] Little, R. J. A. and Rubin, D. B., *Statistical analysis with missing data*. New York: John Wiley and Son, Inc, 1987.
- [23] Molenberghs, G., Michiels, B., Kenward, M. G., and Diggle, P. J., “Monotone missing data and pattern-mixture models,” *Statistica Neerlandica*, vol. 52, no. 2, pp. 153–161, 1998.
- [24] Diggle, P. J. and Kenward, M. G., “Informative dropout in longitudinal data analysis (with discussion),” *Journal of the Royal Statistical Society Series C (applied statistics)*, vol. 43, pp. 49–94, 1994.

- [25] Molenberghs, G., Beunckens, C., Jansen, I., Thijs, H., van Steen, K., Verbeke, G., and Kenward, M. G., *Analysis of incomplete data*. In *Pharmaceutical Statistics with SAS*, Cary, NC: SAS publishing, 2006.
- [26] Spiegelhalter, D. J., Thomas, A., and Best, N. G., *WinBUGS version 1.2 user manual*. Cambridge: MRC Biostatistics Unit, 1999.
- [27] Little, R. J. A., “Pattern-mixture models for multivariate incomplete data,” *Journal of the American Statistical Association*, vol. 88, no. 421, pp. 125–134, 1993.
- [28] White, I. R., Carpenter, J. R., Evans, S., and Schroter, S., “Eliciting and using expert opinions about dropout bias in randomized controlled trials,” *Clinical Trials*, vol. 4, pp. 125–139, 2007.
- [29] Scharfstein, D. O., Daniels, M. J., and Robins, J. M., “Incorporating prior beliefs about selection bias into the analysis of randomized trials with missing outcomes,” *Biostatistics*, vol. 4, no. 4, pp. 495–512, 2003.
- [30] Yan, X., Lee, S., and Li, N., “Missing data handling methods in medical device clinical trials,” *Journal of Biopharmaceutical Statistics*, vol. 19, no. 6, pp. 1085–1098, 2009.
- [31] Mallinckrodt, C. H., Lane, P. W., Schnell, D., Peng, Y., and Mancuso, J., “Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials,” *Drug Information Journal*, vol. 42, pp. 303–319, 2018.
- [32] White, I. R., Kalaitzaki, E., and Thompson, S. G., “Allowing for missing outcome data and incomplete uptake of randomised interventions, with application to an internet-based alcohol trial,” *Statistics in Medicine*, vol. 30, pp. 3192–3207, 2011.
- [33] Gould, A., “A new approach to the analysis of clinical drug trials with withdrawals,” *Biometrics*, vol. 36, pp. 721–27, 1980.
- [34] Lachin, J. M., “Worst-rank score analysis with informatively missing observations in clinical trials,” *Control Clin Trials*, vol. 20, pp. 408–422, 1999.
- [35] Carpenter, J., Kenward, M., Evans, S., and White, I., “Last observation carry-forward and last observation analysis,” *Statistics in Medicine*, vol. 23, no. 20, pp. 3241–3242, 2004.
- [36] Molnar, F. J., Hutton, B., and Fergusson, D., “Does analysis using “last observation carried forward” introduce bias in dementia research?,” *CMAJ*, vol. 179, no. 8, pp. 751–753, 2008.
- [37] Siddiqui, O., Hung, H. M. J., and O’Neill, R., “Mmrm vs. locf: A comprehensive comparison based on simulation study and 25 nda datasets,” *Journal of Biopharmaceutical Statistics*, vol. 19, no. 2, pp. 227–246, 2009. PMID: 19212876.
- [38] Kenward, M. G. and Molenberghs, G., “Last observation carried forward: A crystal ball?,” *Journal of Biopharmaceutical Statistics*, vol. 19, no. 5, pp. 872–888, 2009. PMID: 20183449.
- [39] Molenberghs, G., Thijs, H., Jansen, I., Beunckens, C., Kenward, M. G., Mallinckrodt, C., and Carroll, R. J., “Analyzing incomplete longitudinal clinical trial data,” *Biostatistics*, vol. 5,

- no. 3, p. 445, 2004.
- [40] Hollis, S. and Campbell, F., “What is meant by intention to treat analysis? survey of published randomised controlled trials,” *BMJ*, vol. 319, no. 7211, pp. 670–674, 1999.
- [41] Rubin, D. B., “Multiple imputation after 18+ years,” *Journal of the American Statistical Association*, vol. 91, no. 434, pp. 473–489, 1996.
- [42] Little, R. J. A. and Yau, L., “Intent-to-treat analysis for longitudinal studies with drop-outs,” *Biometrics*, vol. 52, no. 4, pp. 1324–1333, 1996.
- [43] Schafer, J. L., *Analysis of Incomplete Multivariate Data*. Chapman and Hall, 1997.
- [44] Carpenter, J. R. and Kenward, M., *Multiple Imputation and its Application*. Wiley, 2013.
- [45] Schafer, J. L. and Harel, O., “Multiple imputation in two stages,” *In Proceedings of Federal Committee on Statistical Methodology Conference*. Available from https://fcsm.sites.usa.gov/files/2014/05/2003FCSM_Harel.pdf, accessed September 2016, 2003.
- [46] Meng, X. L., “Multiple-imputation inferences with uncongenial sources of input,” *Statistical Science*, vol. 9, no. 4, pp. 538–558, 1994.
- [47] Kenward, M. G. and Carpenter, J. R., “Multiple imputation: current perspectives,” *Statistical Methods in Medical Research*, vol. 16, no. 3, pp. 199–218, 2007.
- [48] Dempster, A. P., Laird, N. M., and Rubin, D. B., “Maximum likelihood from incomplete data via the em algorithm,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.
- [49] Van Buuren, S., “Multiple imputation of discrete and continuous data by fully conditional specification,” *Statistical Methods in Medical Research*, vol. 16, no. 3, pp. 219–242, 2007.
- [50] Azur, M. J., Stuart, E. A., Frangakis, C., and Leaf, P. J., “Multiple imputation by chained equations: what is it and how does it work?,” *International Journal of Methods in Psychiatric Research*, vol. 20, no. 1, pp. 40–49, 2011.
- [51] Hughes, R. A., White, I. R., Seaman, S. R., Carpenter, J. R., Tilling, K., and Sterne, J. A., “Joint modelling rationale for chained equations,” *BMC Medical Research Methodology*, vol. 14, no. 1, pp. 1–10, 2014.
- [52] Zhu, J. and Raghunathan, T. E., “Convergence properties of a sequential regression multiple imputation algorithm,” *Journal of the American Statistical Association*, vol. 110, no. 511, pp. 1112–1124, 2015.
- [53] Schroter, S., Black, N., Evans, S., Carpenter, J. R., Godlee, F., and Smith, R., “Effects of training on quality of peer review: randomised controlled trial,” *BMJ*, vol. 328, no. 7441, p. 673, 2004.

- [54] Busse, W., Chervinsky, P., Condemi, J., Lumry, W. R., Petty, T., and Rennard, S., “Budesonide delivered by turbuhaler is effective in a dose dependent fashion when used in the treatment of adult patients with chronic asthma,” *Journal of Allergy and Clinical Immunology*, vol. 101, pp. 457–463, 1998.
- [55] Cro, S., Morris, T. P., Kenward, M. G., and Carpenter, J. R., “Reference-based sensitivity analysis via multiple imputation for longitudinal trials with protocol deviation,” *Stata Journal*, vol. 16, no. 2, pp. 443–463(21), 2016.
- [56] Cro, S., “Mimix: Stata module to perform reference based multiple imputation for sensitivity analysis of longitudinal clinical trials with protocol deviation,” Available from <https://ideas.repec.org/c/boc/bocode/s457983.html>, accessed 15 July 2016, 2015.
- [57] Carpenter, J. R., Roger, J. H., Cro, S., and Kenward, M. G., “Response to comments by seaman et al. on “analysis of longitudinal trials with protocol deviation: A framework for relevant, accessible assumptions, and inference via multiple imputation,” *Journal of Biopharmaceutical Statistics* 23:1352–1371,” *Journal of Biopharmaceutical Statistics*, vol. 24, no. 6, pp. 1363–1369, 2014.
- [58] Kirkwood, B. and Sterne, J., *Essential Medical Statistics*. Essentials, Wiley, 2003.
- [59] Reiter, J. P., “Multiple imputation when records used for imputation are not used or disseminated for analysis,” *Biometrika*, vol. 95, no. 4, pp. 933–946, 2008.
- [60] O’Kelly, M. and Ratitch, B., *Clinical Trials with Missing Data: A Guide for Practitioners*. Statistics in Practice, Wiley, 2014.
- [61] McMillan, A., Bratton, D. J., Faria, R., Laskawiec-Szkonter, M., Griffin, S., Davies, R. J., Nunn, A. J., Stradling, J. R., Riha, R. L., and Morrell, M. J., “Continuous positive airway pressure in older people with obstructive sleep apnoea syndrome (predict): a 12-month, multicentre, randomised trial,” *The Lancet Respiratory Medicine*, vol. 2, no. 10, pp. 804 – 812, 2014.
- [62] Arden, N., Cro, S., Sheard, S., Doré, C., Bara, A., Tebbs, S., Hunter, D., James, S., Cooper, C., O’Neill, T., Macgregor, A., Birrell, F., and Keen, R., “The effect of vitamin d supplementation on knee osteoarthritis, the video study: a randomised controlled trial,” *Osteoarthritis and Cartilage*, 2016.
- [63] Sterne, J. A. C., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M., and Carpenter, J. R., “Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls,” *BMJ*, vol. 338, 2009.
- [64] Little, R. J. A. ., D’Agostino, R., Cohen, M. L., Dickersin, K., Emerson, S. S., Farrar, J. T., Frangakis, C., Hogan, J. W., Molenberghs, G., Murphy, S. A., Neaton, J. D., Rotnitzky, A., Scharfstein, D., Shih, W. J., Siegel, J. P., and Stern, H., “The prevention and treatment of missing data in clinical trials,” *New England Journal of Medicine*, vol. 367, no. 14, pp. 1355–1360, 2012.
- [65] Pullenayegum, E. M., Platt, R. W., Barwick, M., Feldman, B. M., Offringa, M., and Thabane,

- L., “Knowledge translation in biostatistics: a survey of current practices, preferences, and barriers to the dissemination and uptake of new statistical methods,” *Statistics in Medicine*, vol. 35, no. 6, pp. 805–818, 2016.
- [66] Mackinnon, A., “The use and reporting of multiple imputation in medical research – a review,” *Journal of Internal Medicine*, vol. 268, no. 6, pp. 586–593, 2010.
- [67] Rezvan, H. P., Lee, K. J., and Simpson, J. A., “The rise of multiple imputation: a review of the reporting and implementation of the method in medical research,” *BMC Medical Research Methodology*, vol. 15, p. 30, 2015.
- [68] Royston, P., “Multiple imputation of missing values,” *The Stata Journal*, vol. 4, no. 3, pp. 227–241, 2004.
- [69] StataCorp., “Stata statistical software: Release 11,” *College Station, TX: StataCorp LP*, 2009.
- [70] StataCorp., “Stata statistical software: Release 12,” *College Station, TX: StataCorp LP*, 2011.
- [71] SAS., “Sas version 8.2,” *Cary, NC: SAS Institute Inc.*, 2001.
- [72] SAS., “Sas version 9.3,” *Cary, NC: SAS Institute Inc.*, 2011.
- [73] SPSS., “Spss statistics for windows: version 17.0,” *Chicago:SPSS Inc*, 2008.
- [74] Roger, J. H., “Sas code for sensitivity analysis using multiple imputation,” *Available from www.missingdata.org.uk, accessed 15 July 2016*, 2012.
- [75] StataCorp., “Stata statistical software: Release 14,” *College Station, TX: StataCorp LP*, 2015.
- [76] Lu, K., “An analytic method for the placebo-based pattern- mixture model,” *Statistics in Medicine*, vol. 33, no. 7, pp. 1134–1145, 2014.
- [77] Ayele, B. T., Lipkovich, I., Molenberghs, G., and Mallinckrodt, C. H., “A Multiple-Imputation-Based Approach to Sensitivity Analyses and Effectiveness Assessments in Longitudinal Clinical Trials,” *Journal of Biopharmaceutical statistics*, vol. 24, no. 2, pp. 211–228, 2014.
- [78] Tang, Y., “Short notes on maximum likelihood inference for control-based pattern-mixture models,” *Pharmaceutical Statistics*, vol. 14, no. 5, pp. 395–399, 2015.
- [79] Liu, G. F. and Pang, L., “On analysis of longitudinal clinical trials with missing data using reference-based imputation,” *Journal of Biopharmaceutical Statistics*, vol. 0, no. 0, pp. 1–13, 0.
- [80] Ibrahim, J. G. and Chen, M.-H., “Power prior distributions for regression models,” *Statistical Science*, vol. 15, no. 1, pp. 46–60, 2000.
- [81] Ibrahim, J. G., Chen, M.-H., Gwon, Y., and Chen, F., “The power prior: theory and applications,” *Statistics in Medicine*, vol. 34, no. 28, pp. 3724–3749, 2015.

- [82] Anderson, T. W., “Maximum likelihood estimates for a multivariate normal distribution when some observations are missing,” *Journal of the American Statistical Association*, vol. 52, no. 278, pp. 200–203, 1957.