Causal inference in process evaluation

Development of statistical methods for causal inference in process evaluation of mental health related trials

Li, R

Awarding institution:
King's College London

Causal inference in process evaluation

Development of statistical methods for
causal inference in process evaluation of
mental health related trials

Li, Ruijie

October 2022

A thesis presented for the degree of Doctor of Philosophy

Department of Biostatistics & Health Informatics
School of Mental Health & Psychological Sciences
Institute of Psychiatry, Psychology & Neuroscience
King's College London
United Kingdom

Abstract

Background

Process evaluation (**PE**) focuses on assessing the causal relationships between interventions, outcomes of interests and the processes that bring about changes to the outcomes of interest. Since causal relationships are of interest in such a study, methods of causal inference should generally be applied to **PE**. This thesis focuses on the development of methods of causal inference to overcome existing barriers to the application of these methods in a **PE** with a focus on **PE** in mental health related trials. The focus on mental health related trials is because many of the interventions used in mental health related services affect the outcome via a series of processes and so require the use of **PE** to investigate the causal relationships between the intervention, processes and outcome of interest. The methodological gaps addressed in this thesis were chosen following a systematic review on methods of causal inference used within **PE** in mental health related trials. The gaps in methods of causal inference centre around the estimation of sequentially mediated causal effects and the estimation of the causal odds ratio (**OR**) for binary outcomes.

Overview and main findings

This thesis is broadly divided into four main parts. The first part describes a systematic review of existing methods of causal inference in use for **PE** and identified existing methodology gaps in the use of causal inference within **PE** . The review confirmed the existence of two methodological gaps: estimation of sequentially mediated causal effects with two or more

mediators, and estimation of causal **OR** under scenarios where a continuous mediator is nested within a binary outcome.

The second part of this thesis reviewed the existing methods used to address these problems. From this review the potential outcome (**PO**) framework was adopted to define causal effects of interest within this thesis. This definition would subsequently guide the direction for the development of the methods. Additionally from this review, the estimation method developed by Imai et al. (2010) was chosen to propose novel methodological solutions in this thesis.

The third part of the thesis focused on the development of the methods, focusing on how the foundation of the methods were formed from prior work in causal inference and adaptations to existing methods. A set of procedures was simultaneously developed to validate the new methods. This was done to provide confirmation that the novel methods, as implemented in this thesis, provided results in line with what was expected. The procedures largely verified that these were within expectations. Additionally, sensitivity analyses approaches were proposed for the newly developed methods. To estimate the causal effects using the novel methods, certain assumptions were made and the sensitivity analyses served to assess the impact on the causal effect estimates should the assumptions be violated. A new *R* programme was developed as an implementation of the novel estimation methods as well as the sensitivity analyses and allows these methods to be easily accessible.

The fourth and last part of the thesis consisted of an application of the methods using data from a real randomised controlled trial (**RCT**), the Carers' Assessment, Skills and Information Sharing (**CASIS**) trial. The application demonstrated that the novel methods had real world utility and enabled the testing of hypotheses which previously could not be tested.

Conclusions

This thesis uniquely enables the estimation of sequentially mediated causal effects and the estimation of causal **OR** under scenarios that previously had no existing solutions. Additionally, sensitivity analyses were developed for the newly developed estimators to enable an assessment of assumptions used in the estimation of the causal effects. An important limitation in the current

implementation of the estimators is the inability to include any interaction terms. This can be rectified in a future modification of the developed estimators. An important strength of this thesis lies in the rigorous procedures used to demonstrate the correctness of the implementation of the estimators. Also, a demonstration of the newly developed estimators on a real trial demonstrated their capabilities in testing hypotheses relating to sequentially mediated causal effects and causal **OR** within the conduct of **PE** in an **RCT**.

Acknowledgements

This piece of work came about because many people were willing to give me a chance, a chance to reach greater heights. My head of department at the National Healthcare Group, Singapore, Dr Heng Bee Hoon and my mentor, Prof Teo Yik Ying both of whom gave me invaluable assistance to get me started on my studies. Thanks to Prof Sabine Landau who accepted me as a student despite the lack of formal training in statistics without whom I would not have had the chance to learn from many esteemed colleagues. She also provided me with the guidance needed to build my skills and knowledge in statistics over the many discussions we had. Dr Kimberley Goldsmith, who was my second supervisor was an ever present presence throughout my studies and was always generous with her time, knowledge and kind words. Prof Richard Emsley who later took over as my supervisor was also generous with his encouragements and helpful insights.

Most of all, I need to thank the unwavering support given by my wife, Michelle and my two daughters, Chenxi and Chenling. Thank you for giving me the reason to carry on and for always accommodating. We will, as a family, look forward to better years ahead.

Contents

List of Figures

List of Tables

Abbreviations

ACDE	average causal direct effects
ACME	average causal mediated effect
AESED	Accommodating and Enabling Scale for Eating Disorders
AN	anorexia nervosa
BMI	body mass index
CACE	complier average causal effect
CASIS	Carers' Assessment, Skills and Information Sharing
CSRI	Client Service Receipt Inventory
DASS-21	Depression, Anxiety and Stress Scale Short Form
DE	direct effect
DOR	direct odds ratio
ECHO	Experienced Carers Helping Others
ED	eating disorders
EDE-Q	Eating Disorder Questionnaire
EDSIS	Eating Disorders Symptom Impact Scale
FQ	Family Questionnaire
GLM	generalised linear model
IE	indirect effect
$\mathbf{IE}_1$	indirect effect through M_1 only
$\mathbf{IE}_2$	indirect effect through M_2 only
$\mathbf{IE}_3$	indirect effect through M_1 and M_2
IOR	indirect odds ratio
ITT	intention-to-treat
LHS	left-hand side

MeSH	Medical Subject Headings
MLE	maximum likelihood estimation
MRC	Medical Research Council
MVN	multivariate normal distribution
NCBI	National Center for Biotechnology Information
OE	outcome evaluation
OR	odds ratio
PDF	probability density function
PE	process evaluation
PMF	probability mass function
PNDE	pure natural direct effect
PNIE	pure natural indirect effect
PO	potential outcome
QOL	quality of life
RCM	Rubin Causal Model
RCT	randomised controlled trial
RHS	right-hand side
SCM	structural causal model
SEM	structural equation model
SIGN	Scottish Intercollegiate Guidelines Network
TAU	treatment as usual
TE	total effects
TNDE	total natural direct effect
TNIE	total natural indirect effect
TOR	average treatment odds ratio
vs.	versus
WHOQOL	World Health Organization Quality of Life

Notations

This section serves as a reference to the notations used throughout in thesis.

Equations

1. '$\triangleq$' stands for 'is defined as'.
2. '$:=$' stands for 'is assigned to'. This is used in the context of assigning a variable to the results of a computation.

Variables

Variables used in this thesis belong to one of the following three groups: named variables, latent variables any other variables. The named variables are variables which are given fixed names as follows:

3. 'N' refers to the population size.
4. 'n' refers to the sample size.
5. 'i' refers to a specific individual within the sample (ith person).
6. 'p' refers to the number of independent variables within a model.
    - There are instances where multiple models are being discussed for which the use p is ambiguous.
    - In such instances, the context of p will be clarified at the point of use (e.g. p of the Y model and p of the M model).
7. '$\boldsymbol{X}$' refers to the matrix of sample data.
    - When referring to the subset of sample data used for a model, a subscript indicating the model is used (e.g. $\boldsymbol{X}_Y$ refers to the matrix of sample data used for the model Y).
    - The matrices with a subscript also includes as its first column, a vector of 1s to

represent the intercept. This matrix is also more commonly known as the model matrix.

- Each row of the matrix represents a single subject and each column represents an independent variable so the dimension of the matrix is given by $n \times p$.

8. 'j' refers to a specific independent variable within the model (e.g. the jth variable).
9. 'm' refers to the total number of simulations conducted.
10. 'k' refers to a single, non-specific run of the simulation (e.g. kth run of the simulation). If a specific run required referring to, it will be mentioned as 'for the simulation where $k = 1$' to refer to the first run of the simulation.

Latent variables are used in two contexts within this thesis.

11. The first is in the discussion of binary outcomes modelled using the logistic regression. For a binary outcome Y, the latent variable Y^* is used to represent the logit transformation of the latent quantity $\Pr(Y = 1)$. An asterix is used with the original variable to indicate the logit transformation of the probability of the variable being 1.
12. The second context is in sensitivity analysis where latent variables are used to represent unmeasured confounders. U is used to notate these variables with subscripts to differentiate between multiple U.

The rest of the variables used in this thesis are either vectors or matrices. There are named using the following rules.

13. Vectors and matrices are notated in bold, italic, letters.
    - Vectors use small letters with an arrow on top (e.g. $\vec{\boldsymbol{a}}$).
    - Matrices use capital letters (e.g. $\boldsymbol{A}$).

Functions

Functions refers to mathematical functions and are written with a name or letter followed by round brackets. Enclosed within the brackets are the argument or input of the function. The only exception to this is the *expectation* function which uses square brackets following established conventions. Functions will be defined on first use.

Another function with established use is the I function which is the indicator function. It

evaluates the argument supplied and returns 1 if the evaluation returns true and 0 if it returns false (e.g. $\mathrm{I}(2 > 0)$ will return 1).

Short hand

Some terms are given a shortened form as follows:

14. '**LHS**' and '**RHS**' stands for 'left-hand side' and 'right-hand side' respectively. They are used to refer the left and right hand side of a mathematical expression.
15. 'Pr' is used to represent probability. The probability of $Y = 1$ is written as $\Pr(Y = 1)$.

Subscripts and reference

16. Numeric subscript of vectors and matrices refer to a specific location within the vector and matrices.
17. Non-numeric subscript refer to a non-specific element within the vectors and matrices.
18. Numeric indexes start from 1.
    - $\vec{\boldsymbol{a}}_{\boldsymbol{i}}$ refers to the ith element of $\vec{\boldsymbol{a}}$.
    - $\boldsymbol{A}_{\boldsymbol{i},\boldsymbol{j}}$ refers to the ith and jth element of a two-dimensional matrix.
19. Vectors and matrices surrounded by round brackets refer to versions of them for example in a simulation.
    - $(\vec{\boldsymbol{a}})_k$ refers to the $\vec{\boldsymbol{a}}$ simulated in the kth run of the simulation.

Distributions

20. A normal distribution with mean, μ, and variance, σ^2, is written as $\mathcal{N}(\mu, \sigma^2)$.
21. A multivariate normal distribution (**MVN**) is written similarly with an 'mv' subscript, $\mathcal{N}_{\mathrm{MV}}(\mu, \sigma^2)$. An important difference between the two is that the mean for the **MVN** is a vector and the variance is a variance-covariance matrix.
22. A logistic distribution with mean, μ, and variance σ^2 is written as $\mathcal{L}(\mu, \sigma^2)$.
23. A standard logistic distribution is written as $\mathcal{L}_{std}$ without the mean and variance parameters. The mean and variance of a standard logistic distribution is 0 and $\frac{\pi^2}{3}$ respectively.

Chapter 1

Introduction

Process evaluation (**PE**) is a form of evaluation commonly used in the biomedical domain which examines the relationships between a treatment under study with the outcome of interest. The primary focus as the name suggests is on the processes initiated by the treatment that led to the outcome of interest. The influence of the context within which the treatment was carried out is also of interest in **PE** (Moore et al., 2015a, 2015b). The purpose of such an evaluation is to improve the understanding of the processes that brought about the change in the outcome of interest. This improved understanding can in turn aid in identifying ways to modify the treatment for better outcomes.

Another way to consider **PE** is to contrast it with its better known counterpart, the outcome evaluation (**OE**). **OE** is a form of evaluation that focuses on an outcome or several outcomes of interest. One of the most well-known applications of the **OE** is within the randomised controlled trial (**RCT**). **RCT**s are a type of trial that aim to compare the outcomes between two treatments or more. These treatments being compared typically consist of an experimental treatment and a control treatment. The control treatment often used is the treatment that is currently in practice or treatment as usual (**TAU**). Alternatively, no treatment can be used instead as a control treatment depending on the aim of the trial. The two treatments being compared are randomly offered to the subjects recruited for the study. The aim of such a trial is to be able to assess the relative effectiveness of the novel treatment when compared to the control treatment. More specifically, the question posed by an **RCT** is whether or not an experimental treatment produces better outcome than the control treatment. The outcome in such a trial is one that bears

relevance to the condition that the treatment aims to improve. The randomisation of treatment offered makes it more likely than not that the characteristics of subjects within each treatment group are equally distributed at baseline. The only difference between the treatment groups that remains should then be the treatment offered. This single point of difference allows the researcher to attribute differences in the outcomes between the treatment groups solely to the treatment offered. Through randomisation, the **RCT** avoids the situation where the differences in outcomes can be attributed to differences in the characteristics between the subjects offered the experimental and control treatment. This attribution of differences is an assertion of causality since the presence of a difference in outcome due to a single difference, i.e. the offer of treatment, means that the difference in treatment is what caused the difference in outcome. Subsequent mentions of **RCT** refer to a trial comparing the outcome between a control and an experimental treatment where half the subjects are randomised to each of the treatments. The **OE** is thus a powerful tool to answer a very specific question: which of the two treatments is more effective?

The difference between the **OE** and **PE** lies in the question that each seeks to answer. The **OE** focuses on whether or not the treatment caused a change in the outcome while the **PE** focuses on the processes that the treatment took to change the outcome. Of note is that the **OE** here refers primarily to non-pharmacological treatments where the treatments may be more heavily influenced by factors in the subject's environment. The processes considered in **PE** include how the treatment was implemented as well as the mechanisms of impact. The environment or context under which these processes takes place also play an important role in influencing how the treatment affects the outcome. Under different contexts, the same processes may exhibit different characteristics or different processes may emerge (Moore et al., 2015a, 2015b). Crucially, while the **OE** provides an assessment of whether or not an experimental treatment has better outcomes than the control treatment, it does not provide an indication of how or why the treatment works. An **OE** is typically informed by a theoretical model and the assessment of an experimental treatment being superior than the control treatment can be seen as a validation of the theoretical model. This however may not be well justified in all cases because the theorised mechanisms through which the theoretical model works are not directly assessed. Conversely, should an **OE** demonstrate that the experimental treatment does not produce better outcomes than the control treatment, this again does not indicate that the theorised mechanisms does not exist. The theorised mechanisms may not have produced the expected effects due to the context under which the treatment was conducted. Each of the two

scenarios where the theorised mechanisms of impact are not directly assessed represents a gap in our knowledge of the treatment which the **PE** can fill. The use of **PE** to study the processes and the contexts within which they occur can aid in directly assessing mechanisms of impact hypothesised by the theoretical model. This allows us to gain a better understanding of the treatment beyond whether it produced better outcomes than the control treatment or not. This use of **PE** and its methods to refine our understanding of how and why a treatment works is the focus of this thesis.

The concept of **PE** may be new to some and its relative obscurity compared to the **OE** may lead one to conclude that it is a recently developed form of evaluation. This is not the case as evidenced by mentions of **PE** dating back at least thirty years (Judd, 1987). It was however known then as *process analysis* but it refers to identical concepts as **PE** that we know today. One of the foci of discussion back then was how process analysis could be used in tandem with **OE** for the purposes of causal inferences (Judd, 1987). This thesis continues this discussion with the exploration of current uses of **PE**, the gaps in methods enabling these uses and development of suitable methods to bridge these gaps. The next section focuses on one of these uses: causal inference.

1.1 Conduct of causal inference in PE

This section focuses on the use of **PE** in scientific studies to conduct causal inference. Many of these studies, in their use of **PE**, refer directly to causal relationships between the treatment under study, the outcomes of interest and the processes that bring about those outcomes. The strongest indication that **PE** is of keen interest within the scientific community is from the development of a set of guidelines for the conduct of **PE** in complex interventions by the Medical Research Council (**MRC**) (Moore et al., 2014, 2015a, 2015b) which was subsequently updated (Skivington et al., 2021a, 2021b). The guidelines indicated the use of **PE** in three main areas (Moore et al., 2015a, 2015b). The first being the processes of implementation, the second is the mechanisms of impact and the third is the contextual influences on the implementation and outcomes. While the **MRC** guidelines refer specifically to the use of **PE** in complex interventions, many of the aspects of **PE** are also relevant for other forms of interventions. Its use in complex interventions are especially notable because in complex interventions, there is a pressing need

to understand the interplay between the multiple components of the intervention and how they affect the outcome of interest. **PE** offers a systematic way to understand this interplay and ensures that critical areas of interplay are not overlooked. The use of **PE** in implementation processes investigates whether variations in implementation affect the outcomes of interest. This is especially important when the intervention is complex where large variations of how the intervention is administered could exist in a study. The mechanisms of impact refer to how the intervention changes the outcome of interest. These mechanisms refer to the chain of events initiated by the intervention which eventually ends up modifying the outcome of interest. The intervention could act on the outcome of interest either directly or indirectly through the mechanisms. The pathways through which the intervention affects the outcome are known as causal pathways. The third use of **PE** discussed in the **MRC** guidelines focused on how the context or the environment within which the intervention was administered. In this manner of usage, **PE** can be used to quantify characteristics of the environment and assess whether the context affects the outcome of interest in any meaningful way.

Of the three main uses of **PE** highlighted by the **MRC** guidelines, we can see that each addresses a question of causal inference. The first asks the question 'How does variation of implementation cause a change in the outcome?' while the second asks 'What pathways does the intervention take to cause a change in the outcome?' and the third asks 'What contextual factors causes a change in the outcome?'. Within a causal framework, assessing the changes of measures of implementation in relation to the outcome can be considered as a form of *post-randomisation treatment effect modification* (Dunn et al., 2015). In the context of an **RCT**, this form of causal inference belongs to a class of scenario where a variable measured after randomisation has an effect on the outcome. Data collected on this variable is also notably missing in half of the subjects. Using the subject's adherence to the treatment as an example, if data related to adherence was collected in both arms of the trial, it might, at first glance suggest that we can compare the adherence between the two groups directly. The is however not the case because adherence is specific to the treatment, i.e. adherence to the control treatment is a different variable from adherence to the experimental treatment. Therefore, for the subjects who were offered the control treatment, adherence can only be observed for the control treatment for this group of subjects. Vice versa is also true for subjects who were offered the experimental treatment. Hence, if we were interested in the effect of adherence to the experimental treatment on the outcome, we would be missing data for adherence to the experimental treatment for

the subjects offered the control treatment (Dunn et al., 2015). The missing data in this scenario is important because without information on the adherence to the experimental treatment for subjects offered the control treatment, the outcomes observed for the subjects offered the experimental treatment would be confounded by the adherence and the treatment offered, i.e. we would not be able to separate out how much of the change in the outcome was due to the treatment and how much was due to the better or poorer adherence. The third question addresses a question quite similar to post-randomisation treatment effect modification. In the third question, the researcher is interested in which contextual factors affect the outcomes and to what degree. In this scenario, if the contextual factors of interest are treatment independent and are present before randomisation, then the estimation of the effects of the contextual factors on the outcomes belong to a class of scenarios known as *moderation* (Baron & Kenny, 1986; Edwards & Lambert, 2007). Should the contextual factors be present only after randomisation, then it has a similar missing data problem for the post-randomisation treatment effect modification scenarios and should be addressed similarly. The second question on mechanisms of action focuses on the pathways through which the causal effects occur. This class of scenario is known as *mediation* (Baron & Kenny, 1986; Dunn et al., 2015; MacKinnon, 2008) and forms the primary focus of this thesis.

Much like **PE**, *mediation* or *mediational analysis* dates back at least for decades and one of the better known pieces of work on mediational analysis comes from Baron and Kenny (1986). The work by Baron and Kenny (1986) sought to clarify the distinction between mediators and moderators as well as highlighting considerations in estimating mediated and moderated causal effects. More importantly the assessment of mediated and moderated effects always takes place in the context of assessing causality. This emphasis on causality, when considered together with the envisioned use of **PE** in understanding mechanisms of action indicates that mediational analysis can be a tool to conduct such a study into the mechanisms. However, the study of causal mechanisms additionally requires the question to be better defined and the same can be said of the other causal inference questions that **PE** seeks to address. A question such as 'Through which pathways does the intervention work to affect the outcome?' is far too general and there needs to be a process of refinement of this question. This refinement could rely on existing theoretical knowledge, a point heavily emphasised in the **MRC** guidelines on the conduct of **PE**. Theoretical knowledge could indicate the pathways an intervention is likely to take to affect the outcome of interest and **PE** could be used to test such a hypothesis. There could also

be pathways that have not been formally theorised but may already have been considered as possible candidates of intermediaries that an intervention goes through to act on the outcome. Notably, the **MRC** guidelines highlighted the use of qualitative methods as a means to gather such information. Qualitative methods are a set of methods that use non-quantitative data such as interviews, focus groups and observations by the researcher of various aspects of the intervention to answer specific questions. In this instance, qualitative methods can be used to gather varied as well as in-depth perspectives of the various processes of the intervention from different people involved in the intervention including the patients receiving, and the people administering the intervention. The information gathered can be used to uncover previously unknown processes within each of the three questions posed by **PE**. This provides a systematic way to generate hypotheses which in turn aids to refine questions that can then be answered using methods of causal inference through an appropriately designed **PE**.

The use of causal inference within **PE**, as discussed in the **MRC** guidelines, has the potential to provide valuable insight into how interventions work. These insights can be used to modify the interventions for specific aims such as better outcomes, lower costs or easier implementation. Knowing precisely the relationship of the processes within an intervention with its outcome allows similarly precise ways to modify the intervention. These insights, however, depend on the ability to conduct causal inference within **PE** and there are two notable barriers.

The first concerns the definition of causal effects, specifically mediated causal effects. Mediated causal effects are effects that go through an intermediary to act on the outcome. More formally, a mediator is a variable that lies on the causal pathway between the intervention and the outcome. Changes in the independent variable changes the mediator and the mediator in turn accounts for variations in the outcome of interest (Baron & Kenny, 1986). An effect that goes through a mediator is said to be an *indirect effect (IE)* and an effect that does not is said to be a *direct effect (DE)*. The effect of the treatment on the outcome of interest regardless of the path it takes is known as the *total effects (TE)*. The definition for **TE** is well defined. Consider a typical **RCT** comparing an outcome of interest between a control and an experimental treatment. The **TE** is defined as the difference between the outcomes of the groups offered the experimental and control treatment, which is also the main effect of interest in a typical **RCT**.

IEs however were less well-defined. This was the case until the introduction of potential

outcome (**PO**) to aid in defining **IE** (Pearl, 2001). **PO** are simply outcomes that could happen. This possibility of an outcome occurring gives rise to the use of the word 'potential'. Consider a binary outcome of interest in an **RCT** with values of 0 and 1. For a given treatment, the outcome for any subject in the **RCT** could be 0 or 1. The outcome would only be known after it had been observed but before this observation had been made, both of the outcomes had the potential of being realised.

Prior to the use of **PO** to define **IE**, the definition of **IE** had been tightly coupled with the methods used to estimate it (e.g. see MacKinnon (2008)). Two of the commonly used methods to estimate the **IE** are known as *product of coefficients* and *difference in coefficients* methods (MacKinnon, 2008). Each of these methods convey a different interpretation of what an **IE** is and these methods were initially only applied to estimate mediate causal effects when both the outcome and mediator are both continuous. The product of coefficients method defines the **IE** as the amount of change in the dependent variable due to a unit change in the independent variable indirectly through a mediator. The difference in coefficients method defines **IE** as effects that are not direct. The two definitions refer to the same quantity under many but not all scenarios. Specifically, when the regression model used to estimate the **IE** is not linear such as the case in a logistic regression, the two methods of estimation produced different causal **IE** estimates. This creates a situation where estimates of the **IE** can be different depending on how it is estimated and it is not clear when each method is appropriate for use. This lack of a causal effect definition independent from the estimation methods made adaptation of the estimation of mediated causal effects beyond the simplest scenarios difficult (Pearl, 2009). This changed with the use of the **PO** framework to define causal effects and notably the use of this definition to create a general way in which mediated causal effects can be defined known as the *mediation formula* (Pearl, 2001). The use of **PO** allows for hypothetical constructs that aid in the definition of **IE**s including constructs that would not be directly observed within the trial and constructs that could not exist in reality. Despite the fact that some of these constructs would not be observed within a trial and some of these constructs could not possibly exist in reality, they provide an essential crutch in defining **IE**s. The **PO** framework, in its simplest form, considers what would have happened had a preceding event been different? Consider an observation that after taking a pill, a person recovers from his illness. We can use the **PO** framework to consider the outcomes that could have happened had the pill been taken and had the pill not been taken. A comparison of these outcomes under certain circumstances, such as in the case of an **RCT**,

allow us to make causal inferences about the effect of the pill on the illness. This breakthrough in the definition of causal effects subsequently led to the development of a set of formulae known as the *mediation formulae* which uses the aforementioned hypothetical constructs. While they are called formulae, these formulae are actually formal definitions of mediated causal effects that are independent of any estimation methods. They express the mediated causal effects using **PO** and from the expressed **PO**, map them onto statistical concepts such as expectations and probabilities. They do not however state how the expectations and probabilities are to be estimated and allow the researcher to use any method deemed suitable and fit for their specific purpose. The development of the mediation formula also paved the way for many new methods used to estimate mediated causal effects of which the work in this thesis is one of them.

The success of the mediation formula rested on two main pillars. Firstly, the **RCT**, an accepted experimental design for assessing causal relationships, can be framed as a comparison of **PO** (Hernan & Robins, 2020; T. J. VanderWeele & Vansteelandt, 2009) and such a framing of a well-understood experimental design lends credence to the **PO** framework. The second important development was the long history of the discourse of **PO**, notably the work by Lewis (2001) where the logic of **PO** was discussed extensively. This provided the foundation for some of the assumptions of defining causal effects using **PO**. This foundation was critical because causality had been studied since ancient times and any definition of causal effects cannot ignore the large amount of prior work. The work by Lewis (2001) ensured that the use of **PO** in defining causal effects takes into account and is consistent with prior work. These two pillars thus provided the foundation upon which the mediation formula was built on. An area which the original mediation formula had not addressed was the issue of defining causal effects for multiple mediators. Daniel et al. (2015) notably enumerated all the possible sets of definitions of direct and indirect effects in the presence of multiple mediators and noted the high level of complexity as the number of mediators grew. Daniel et al. (2015) further proposed ways by which this complexity be reduced by focusing only on paths that are of interest, i.e. paths that are present within a theoretical model.

Having a definition for causal effects, the attention next turned to how the causal effects could be estimated. With the use of the mediation formula, the estimation became more straightforward since one simply has to estimate the quantities found within the mediation formula. However, in the definitions of mediated causal effects are **PO** nested within another **PO** and this made

estimation of some mediated causal effects difficult. This is the case when the outcome is binary and modelled using a logistic generalised linear model (**GLM**) while the mediator is continuous and modelled using a normal **GLM**. Both the logistic and normal **GLM** belong to a family of linear models that seek to generalise different linear models under a single framework (McCullagh & Nelder, 1989). The problem with the nested **PO** was noted by Valeri and VanderWeele (2013) who also devised a solution. This solution however is an approximation of the causal effect and this approximation is valid only when the event rate of the outcome is rare. This limited the utility of the solution and more importantly, despite having well-defined mediated causal effects, a general method of estimation was still out of reach. This was also true of the case with multiple mediators.

1.2 Aims and objectives

From the previous sections, I had laid out the potential of **PE** in aiding our understanding of the inner workings of an intervention. This understanding requires the use of causal inference methods which at present have certain limitations. The two limitations highlighted are the lack of methods for estimation of causal odds ratio (**OR**) when a continuous mediator is nested within a binary outcome and the estimation of sequentially mediated causal effects in the presence of multiple mediators. These limitations imply that should questions of mediation be of interest in a **PE**, they would be absent from current literature or addressed in ways that avoids the use of methods of causal inference because the estimation methods do not exist. This thesis sets out to overcome these two limitations.

The first aim (chapter 2) is to confirm that questions of causality are of interest in **PE**. This is done by conducting a systematic review of studies using **PE** with a focus on mental health related trials. The focus on mental health related trials is due to the nature of the interventions involved which often include indirect mechanisms of action between the intervention and outcome. Such indirect mechanisms make the practice of **PE** highly relevant and hence the focus on mental health related trials. Given the barriers highlighted, it is predicted that studies with an aim of assessing causality within **PE** may address them in ways that avoids the barriers. By examining the stated aims of conducting **PE** as well as the methods used to achieve those aims, we can assess whether the inference of causality is one of the aims and whether or not a gap in methods

exist.

The second aim (chapter 3) is to lay out how the **PO** can be used to address the two gaps identified. This aim is achieved by examining how the **PO** framework had been used to define causal effects and mediated causal effects. A review of existing methods for estimation is also done to determine how each of the existing methods have addressed various issues in the estimation of mediated causal effects.

The third aim (chapter 4) is to first develop a definition for the causal **OR** and sequentially mediated causal effects and then develop estimating methods for each of them. The development of the definitions is done by examining the structure of the mediation formula and using the method used to construct the mediation formula, to extend it to derive definitions for the causal **OR** and for sequentially mediated causal effects. Estimation methods are then developed, drawing upon existing methods.

The fourth aim (chapter 5) is to verify that the novel estimators developed in chapter 4 were 'within expectation'. This first required a definition of 'within expectation'. The choice adopted differs from that used in traditional methods of validation of estimators and the rationale for this deviation is explained in the chapter. An application of novel estimators on a real **RCT** was also conducted to demonstrate the utility of the methods.

The fifth aim (chapter 6) is to address the assumptions inherent in the estimation of the causal estimands. Given that it is not possible to determine if these assumptions were violated, methods were instead developed to determine the degree to which the outcomes would change should the assumptions be violated. This is known as the sensitivity analysis. An application of the sensitivity analysis is also conducted to demonstrate its utility in a real world trial.

The last chapter (chapter 7) in this thesis reviews the scope of the work that had been done and highlights the unique contributions of this thesis. These unique contributions go beyond their use in this thesis and have wider applicability in methods of causal inference. The strengths and weaknesses of this thesis are also discussed.

Chapter 2

Review of methods used for process evaluation in mental health related trials

2.1 Introduction

This chapter provides a review of the methods traditionally used to conduct **PE**. The aim is to provide an overview of current methodological practice in **PE** and to identify methodological areas in need of better application of existing methods or further methodological development. A systematic review of recent published studies of **PE** was conducted for this purpose. Areas of **PE** where new methods are needed will be discussed at the end of the chapter. The development of the said methods would be addressed in subsequent chapters.

The central task of this systematic review is to identify the research questions posed by **PE** and the corresponding methods used to answer those questions. The systematic review was limited to **PE** conducted alongside a randomised controlled trial (**RCT**) within the mental-health domain. The focus on mental-health related trials was motivated by the prevalent use of **PE** in such studies. In addition, randomisation in trials simplifies the outcome evaluation (**OE**) and thus such study designs are a good starting point to understand the type of methods used in **PE**. Specifically, randomisation should avoid any confounding of the effect by treatment assignment on the outcome of interest and so enable the estimation of the causal effect of the treatment under investigation using simple methods. Process evaluation can then be used to try

and understand the processes that lead to these causal treatment effects.

The review of studies reporting process evaluations in trials focused on four aspects of **PE**: stated aims, research design, analytical methods and stated conclusions. These four aspects provide information on the motivation, the main problem and the approach to the problem posed by the **PE**. First, the stated aims convey the processes that are being evaluated and the research questions that these aims address. Second, the research design and analytical methods lays out the data were collected and analysed to generate evidence in support of the stated aims. Third and final, the stated conclusions show how the generated evidence had been used in support of the stated aims. More pertinent to our purpose, these four aspects of a **PE** study provide insight into methodological needs of the researchers driven by the problems to be answered by the **PE**. From these methodological needs, gaps can then be identified and addressed in subsequent chapters.

Prior to moving on to the methods used to conduct the systematic review, it is important to highlight the differences between **PE** and **OE**. Outcome evaluation is the main form of evaluation conducted in **RCT**s. The primary aim of such an evaluation is to compare the therapeutic effectiveness or efficacy of an experimental treatment with that of a control treatment. The control treatment is usually a placebo or treatment as usual (**TAU**). The therapeutic effects are measured by predetermined outcome measures that are relevant to the condition for which the treatments are being used for.

PE on the other hand, as the name suggests, is an evaluation of processes. The Oxford dictionary of English defines processes as 'a series of actions or steps taken in order to achieve a particular end' (Soanes & Stevenson, 2005). **PE** can thus be understood as an evaluation of the series of actions carried out to achieve an intended end. In the context of a **PE** of an **RCT**, this intended end is to measure the outcomes of interest on the subjects after the treatments, both experimental and control, have been delivered. The series of actions are thus those that are related to the conduct of the **RCT** such as recruitment of subjects, offer of treatments, delivery of treatments, subjects' adherence and response towards treatment etc. These actions constitute the 'processes' being evaluated within a **PE**.

In a recent Medical Research Council (**MRC**) guideline on **PE** for complex health interventions (Moore et al., 2014, 2015a, 2015b), the value of such an evaluation was highlighted focusing on

three groups of people: policy-makers, practitioners, and researchers. For a policy maker, **PE** aids in identifying contextual factors that makes a treatment more effective. This allows the policy maker to consider what support needs to be in place to create a conducive environment for the treatment to be effective. For a practitioner/clinician, it informs them on potential ways to modify the treatment for better therapeutic benefits which they can adopt in their daily practice. Lastly, for the researcher, results from **PE** allows for a deeper understanding of the interplay of factors that influence the desired outcomes, paving the way to a better understanding of precisely how and why a treatment works.

The examples stated highlights the role **PE** plays in understanding and improving the treatment. Thus, **PE** can be thought of as an evaluation to answer the questions of 'how, why, and for whom does the intervention work?'. This complements the aim of **OE** well, which is to answer the question: 'is the experimental treatment more effective/efficacious than the control treatment?'. Two questions remain however and that is: 'how and why is **PE** being used in practice?' and 'what methodological gaps there are, if any, in meeting the aims of **PE**?'. The first question addresses the motivations behind the use of **PE** in practice and the kind of questions that researchers seek to answer through the use of **PE**. Some of the question that can be answered by **PE** are well placed to tap on advances from the causal inference domain to answer the questions in a principled and rigorous manner. Being able to identify the questions of interest in **PE** that can benefit from the use of causal inference methods is the focus of the second question. The second question focuses on questions that are of a causal nature and whether methodological improvements can be made in the conduct of **PE**. With this, we will move on to the methods used in this review.

2.2 Methods

This section discusses the methods used in the systematic review. It is important to note prior to introducing the methods that this systematic review differs from traditional systematic reviews in important ways. In traditional systematic reviews within the biomedical domain, the focus is on summarising the effects of a treatment of interest within a specific diagnostic group.

These reviews aim to provide an assessment of whether the evidence found in existing literature

supports a specific hypothesis about the treatment, which is typically that the treatment is beneficial to a specific diagnostic group of patients. In order to obtain a trustworthy assessment of the evidence of the hypothesis, the evidence this assessment draws upon needs to be trustworthy in the first place. This is why in traditional systematic reviews, there is a strong focus on assessing the trustworthiness of each study and how well each of them had adhered to best practices in the research that they are conducting. However, in the current review, the aim is not to assess the support for a specific hypothesis. Rather, the aim is to identify the aims and corresponding methods currently used in **PE**. Since the primary motivation for conducting an assessment of trustworthiness is absent and that trustworthiness of a study was deemed to have little bearing on the validity of the stated aims and methods of **PE**, a detailed assessment of trustworthiness was not conducted. A brief assessment using only whether or not the trial was registered with a trial registry and the absence or presence of blinding was used. The purpose of this is to provide a crude gauge of the quality of the studies that the review draws its conclusion upon. Given the unique nature of this review, researchers have taken to calling such reviews a 'scoping review' (Munn et al., 2018) but since the term is relatively new at the time of writing, I shall refer to this review as a 'systematic review'.

The steps taken for the systematic review are listed in table 2.1. Each of the steps are covered in greater detail in the following subsections.

Table 2.1: Steps of systematic review

1. Develop and execute a search strategy.
2. Retrieve and screen the results from the search for inclusion into the review.
3. Extract, analyse and summarise relevant information from the identified studies.
4. Verification with second reviewer that the screening and extraction steps are reproducible.

2.2.1 Conducting the search

A search strategy was required to conduct the search. There were four main parts to the strategy. First, the inclusion criteria for identifying relevant studies needed to be developed. The criteria can be found in table 2.2. Studies not fulfilling these criteria were excluded. Second, databases

Table 2.2: Inclusion criteria for systematic review

1. Study attributes:
   (a) The study must report on the result of a **PE** and must explicitly state so.
   (b) The study must be conducted alongside or after an **OE** aimed at evaluating a treatment for the target population below.
   (c) The study must be from a peer-reviewed journal.
2. Study design:
   (a) The **OE** must adopt an **RCT** design.
   (b) There is no restriction on the study design of the **PE**.
   (c) The primary outcome of the **OE** must be a measure of symptoms associated with a condition that had been defined in the Diagnostic and Statistical Manual of Mental Disorders V (**DSM-V**) (e.g. Patient Health Questionnaire (**PHQ-9**) scores as a measure of symptoms associated with depression which had been defined in the **DSM-V**). A secondary outcome of such a measure does not fulfil the inclusion criteria.
3. Target population:
   (a) No restrictions on the target population of the **OE** and **PE**. It must be noted however that a target population of people with mental health disorders but no mental health related primary outcome (see study design) does not fulfil the study design inclusion criteria (e.g. Vocational rehabilitation for persons with schizophrenia).

and search engines to be used in the search had to be identified. A database holds the metadata of the studies it indexes while a search engine provides the interface to conduct searches of the

metadata within the database. The decision of which databases and search engines were to be used to identify relevant studies was made based upon the coverage and characteristics of each database. The identified databases, search engines and their justifications are listed in table 2.3. Some databases can be searched using multiple search engines and in those circumstances, the search was repeated for each search engine to improve the chances of identifying relevant studies.

Thirdly, suitable keywords had to be identified to be used to execute the search. A close examination of the inclusion criteria and the capabilities of identified search engines indicated that the only portion of the inclusion criteria that can be translated into a search terms for input into the search engines is the requirement for the study to report on the results of a **PE** and for the **PE** to be associated with an **RCT**.

In order to identify studies that report on the results of a **PE**, suitable search terms need to be identified. The words 'process' and 'evaluation' and the various permutations of the two words and their variations were used as search terms but it was unclear if the two words alone were sufficient in identifying all the relevant studies. An indication that it was not sufficient came from a study which made reference to the term 'process analysis' (Judd, 1987; Judd & Kenny, 1981), a concept similar to **PE**. This led to a keyword study where the aim is to identify the different ways of referring to **PE** in recent literature. The details of the keyword study can be found in Appendix A and in order not to detract from the focus of this chapter, the details of the keyword study will not be discussed. The keyword study yielded some alternative ways of referring to **PE** but these were quite similar to the concepts expressed by the words 'process' and 'evaluation'.

The other set of search terms that was needed were the keywords identifying studies conducted with an **RCT** design. This task was made easier by the Cochrane Collaboration and Scottish Intercollegiate Guidelines Network (**SIGN**) as both organisations had previously published specific search strategies to identify studies that make use of the **RCT** design (Higgins et al., 2008; Scottish Intercollegiate Guidelines Network, 2015; Scottish Intercollegiate Guidelines Network (SIGN), 2015). These search strategies were however outdated by the time the search was conducted as there have been major changes to the databases and search engines. The most significant change was the migration of existing Medical Subject Headings (MeSH) terms to a

new iteration of MeSH. While the published search strategies could not be used as they were, they provided important guidance on how to develop a search strategy for studies that use the **RCT** design.

The published search strategy was then modified to adapt to the changes of search engines and databases and in the process of adaptation, decisions needed to be taken. These decisions were typically a choice between potentially excluding relevant studies or potentially including irrelevant studies. The choice taken was always in favour of potentially including irrelevant studies because each study was eventually checked manually to confirm its eligibility for inclusion. This process ensured that the widest net was cast to identify relevant studies although this was at the expense of creating a larger workload to remove ineligible studies. The full search strategy is documented in Appendix B. Lastly, the period to search for studies was limited to the past five years at the time of conducting the search (2012 - 2016). This ensured that the identified studies were an accurate reflection of current practices of **PE**. The publication date was used to identify the date for each study. In all the databases searched, this was a stable field that was not subjected to future retrospective changes. This provided some safeguards that the results presented in this thesis can be reproduced should there be a need to verify or update them.

The search was conducted using the keywords identified and the date ranges specified. The metadata of the studies identified from this search was extracted and each study was screened manually to identify studies that met the inclusion criteria. In most of the cases, the metadata alone (including the abstract) had insufficient information as to whether the inclusion criteria were met. This meant that the fulltext of the study had to be retrieved to assess whether or not to include the study into the review. In about a quarter of the cases, the **OE** associated with the **PE** had also to be retrieved to make the assessment. Given the large numbers of papers that had to be retrieved, an automated system was built to facilitate this task. At the end of this screening process, a set of studies that met the inclusion criteria were identified.

2.2.2 Data extraction and analysis

The first step to data extraction was to retrieve the paper associated with the identified studies. The paper of the **OE** associated with the **PE** was also retrieved. The retrieval used the automated

Table 2.3: Databases searched

1. **CINAHL**
    - Search engine: EBSCOhost
    - Justification: Indexed nursing related literature that were not covered by MEDLINE.
2. Cochrane Central
    - Search engine: Wiley
    - Justification: Had a unique feature where studies using the **RCT** design were identified as they were being indexed by the database, allowing for searches to operate only on the subset of **RCT** studies.
3. Embase
    - Search engine: Embase.com, Ovid
    - Justification: Had a distinct set of journals not covered by the other databases.
4. PsycINFO
    - Search engine: Ovid
    - Justification: Indexed studies related to mental health.
5. PubMed / MEDLINE
    - Search engine: **NCBI**, Embase.com, EBSCOHost, Ovid
    - Justification: One of the widest coverage of biomedical journals.

system discussed in the previous sub-section. For the review, specific pieces of information were extracted from each study and entered into a spreadsheet. Each field within the spreadsheet had a precise definition of the type of information that was required. This is documented in Appendix C, which provides the manual intended for a second reviewer to reproduce the screening and data extraction process. The role of the second reviewer will be discussed in section 2.2.3. A list of the information extracted from the **OE** and **PE** studies are in table 2.4 and 2.5. The

Table 2.4: Information extracted from **OE** studies

1. Trial registration and blinding
    - Provides an assessment of the quality of the trials.
2. **RCT** design, number of trial arms and type of randomisation
    - Describes the **RCT** designs used by the studies.
3. Aim of trial and trial outcome
    - Used to determine the context under which **PE** was conducted.
    - 'Aim of trial' could either be 'efficacy' or 'effectiveness'.
        - The aim stated in the spreadsheet was what the author stated the aim to be. There were no further assessments made on the appropriateness of the stated aim.
    - 'Trial outcome' could either be 'positive' or 'negative'.
        - A positive trial outcome refers to a trial with a conclusion that the experimental treatment had better outcomes than the control treatment and a negative trial outcome refers to a conclusion that the experimental treatment is no better or worse than the control treatment.

information extracted from both the **OE** and **PE** studies was summarised using frequencies and percentages to provide an understanding of the context under which **PE** was conducted, the processes that were of interests and the methods used to accomplish the stated aims.

Table 2.5: Information extracted from **PE** studies

1. Methods of inquiry
    - Used to identify methodological preferences in **PE**.
    - This field could either be 'qualitative', 'quantitative' or 'mixed methods'.
2. Causal inference as aim
    - Provides an assessment of whether or not methods of causal inference were needed within **PE**.
    - This field could either be 'no', 'yes, stated' or 'yes, implied'.
    - The classification of stated or implied depends on whether or not the authors were explicit in stating causal inference as their aim. If the authors stated that the aim of the **PE** was to establish causality either among processes or between processes and outcomes, this was classified as a stated aim. When there was no explicit stated aim of causal inference but the analysis conducted and discussions indicated otherwise, it was classified as an implied aim of causal inference.
3. Processes investigated
    - Provides an understanding of what types of processes were of interest to the community of mental health researchers.
4. Linkage of **PE** results to **OE**
    - Provides an indication of how researchers were making use of the two forms of evaluation, whether in isolation or in tandem.
    - This field can either be 'yes' or 'no'.
    - Any discussion of the results of the **PE** in the context of the **OE** is considered to be a 'yes'. Otherwise, this is a 'no'.

2.2.3 Secondary review by independent researcher

An independent second reviewer was brought in to repeat the conduct of parts of the systematic review. The aim is to ensure that the conduct of the systematic review was documented unambiguously and if one followed this documentation, he or she will be able to reproduce the results presented. This documentation was written as a manual for the second reviewer (Appendix C). The secondary review did not repeat all parts of the systematic review. Only the parts that required some element of judgement was repeated. These were parts that where biases can be introduced and the secondary reviewer serves as a safeguard against this. These parts were the screening of the studies and the extraction of the information into the database. The reviewer was not asked to repeat the entire process because that would have been overly laborious. A random subset of studies (details below) was used for these reproductions and these were considered to be a sufficient gauge of the unambiguity of the conduct of the systematic review. The part that was not repeated was the searching of the databases as there was little chance that this process was ambiguous or had biases introduced since the full search parameters were documented. This secondary review serves as another safeguard to ensure reproducibility and a check to ensure that biases had not been introduced in the conduct of the review.

For the screening of studies, a random selection of ten studies that were included in the review and ten studies that were excluded were given without classification to the reviewer. He was then tasked to follow the manual in which the inclusion criteria were stated to identify studies that were to be included. If any of the studies were classified differently by the second reviewer, this was investigated to determine the cause of it. In any case, any studies that were classified wrongly indicated that there was ambiguity either because the reviewer understood the instructions differently from what was intended or that the instructions were not detailed enough and left room for interpretations.

For the extraction of information into the database, a similar process was used where the reviewer was tasked to extract information from ten studies into the database. Any discrepancies between what was entered by the review and myself was investigated. Specifically, the reviewer and myself discussed the discrepancies and identified the source of the discrepancy. Discrepancies can point either to an ambiguity in the documentation or a mistake in applying the protocol. The former was rectified by modifying the documentation with clearer instructions. This process

reduced the chance that the protocol for the review was interpreted differently from what was intended.

2.3 Results

This section discusses the results from the review by the independent researcher and the systematic review. The screening process repeated by the independent researcher yielded a single discrepancy between what I had included for the review and what the independent researcher had identified as fitting the inclusion criteria. This was found to be due to an ambiguous wording of what constitutes a mental health related trial in the reviewer's manual. This had since been rectified. The data extraction process did not yield any discrepancies. Given the lack of major discrepancies between the results produced by the independent researcher and myself, this indicated that the parts of this systematic review most prone to errors have been mitigated well.

Moving on to the results from the review, the full list of identified **PE** studies can be found in Appendix E. The list of associated **OE** studies can be found Appendix F. Information from the studies had been summarised in table 2.7, table 2.9 and table 2.10. After the initial search and prior to any processing, 3010 results were returned. The number of results by database and search engine can be found in table 2.6. As mentioned previously, some databases were searched using multiple search engines and this meant that a large number of duplicated results was to be expected. After removing duplicated results, 930 articles remained which underwent further screening as discussed at the end of section 2.2.1. After going through a manual screening procedure for each of the 930 articles, 30 articles remained. This large drop in the number of articles was also expected as the search strategy prioritised having larger numbers of false positives over missing relevant articles altogether. The 30 articles remaining reported on the results of 25 unique studies. Some studies reported different aspects of the **PE** in separate articles.

The 25 studies came from 7 different countries with the United Kingdom comprising the largest share of 48% and the Netherlands coming in second with 20% of the studies. Table 2.7 lists the breakdown by each country. It is unclear why the United Kingdom has such a large number of studies using **PE** compared to the other countries. A plausible reason could be that the use of

Table 2.6: Database, search engines and search results count

Database	Search engine	Results count
Pubmed	**NCBI**	578
MEDLINE	Ovid	569
Embase & MEDLINE	Embase.com	426
Embase	Ovid	168
CINAHL & MEDLINE	EBSCOhost	577
PsycINFO	Ovid	404
Central	Wiley	288
	Total	3010
	Deduplicated	930
	Accepted	30
	Rejected	900

PE was driven by requirements or preferences of the different funding agencies within each country and the presence of the **MRC** guidelines on **PE** is an indication that this might be the case. Information on the trial registration, outcome and type of trial can be found intable 2.8.

Table 2.7: Countries of origin of identified studies

Countries	Count	%
United Kingdom	12	48%
Netherlands	6	24%
Australia	2	8%
China	2	8%
Singapore	1	4%
South Africa	1	4%
United States of America	1	4%

Detailed information regarding the trial can be found in Appendix D.

Table 2.8: Accepted studies trial information

No.	Study	Main trial	Process evaluation	Trial registration	Trial outcome	Type of trial
01.	01.	Chan et al., 2011	Chan et al., 2012	none	positive	efficacy
02.	02.	Roy-Byrne et al., 2010	Curran et al., 2012	none	positive	effectiveness
03.	03.	Gao et al., 2010	Gao et al., 2012	none	positive	efficacy
04.	04.	Leontjevas et al., 2013	Leontjevas et al., 2012	NTR1477	positive	effectiveness
05.	05.	Gärtner et al., 2013	Ketelaar et al., 2013	NTR2786	positive	effectiveness
06.	06.	Crawford et al., 2012	Patterson et al., 2013	ISRCTN46150447	negative	efficacy
07.	07.	Stallard et al., 2013	Stallard et al., 2013	ISRCTN19083628	negative	effectiveness
07.	08.		Taylor et al., 2014	ISRCTN19083628	negative	effectiveness
08.	09.	Thomas, Walker et al., 2013	Thomas, Russell et al., 2013	ISRCTN56078830	positive	efficacy
09.	10.	van der Krieke et al., 2013	van der Krieke et al., 2013	NTR3105	negative	efficacy
10.	11.	Arends, Klink et al., 2014	Arends, Bültmann et al., 2014	NTR1963	positive	effectiveness
11.	12.	Underwood et al., 2013	Ellard et al., 2014	ISRCTN43769277	negative	effectiveness
12.	13.	Parry et al., 2016	Finch et al., 2014	ISRCTN78396615	positive	effectiveness
13.	14.	Geraedts, Kleiboer et al., 2014	Geraedts, Kleiboer et al., 2014	NTR2993	negative	effectiveness
14.	15.	Hind et al., 2014	Hind et al., 2014	ISRCTN28645428	negative	effectiveness
15.	16.	Slade et al., 2015	Leamy et al., 2014	ISRCTN02507940	negative	effectiveness
15.	17.		Wallace et al., 2016	ISRCTN02507940	negative	effectiveness

Table 2.8: Accepted studies trial information (cont'd)

No.	Study	Main trial	Process evaluation	Trial registration	Trial outcome	Type of trial
16.	18.	Schmidt et al., 2015	Lose et al., 2014	ISRCTN67720902	negative	efficacy
16.	19.		Waterman-Collins et al., 2014	ISRCTN67720902	negative	efficacy
16.	20.		Zainal et al., 2016	ISRCTN67720902	negative	efficacy
17.	21.	McCann et al., 2013	McCann and Lubman, 2014	ACTRN12609000064202	positive	effectiveness
18.	22.	Petersen et al., 2014	Petersen et al., 2014	none	positive	effectiveness
19.	23.	Prick et al., 2016	Prick et al., 2014	NTR1802	negative	effectiveness
20.	24.	Woolhouse et al., 2014	Woolhouse et al., 2014	ACTRN12613000742774	positive	effectiveness
21.	25.	Foster et al., 2016	Myall et al., 2015	ISRCTN67521059	negative	effectiveness
21.	26.		Foster et al., 2016	ISRCTN67521059	negative	effectiveness
22.	27.	Shorey, Chan et al., 2015	Shorey, Chan et al., 2015	ISRCTN15886353	positive	effectiveness
23.	28.	Stallard et al., 2015	Stallard et al., 2015	ISRCTN23563048	positive	effectiveness
24.	29.	Sayal et al., 2016	Taylor et al., 2015	ISRCTN87634685	negative	effectiveness
25.	30.	Priebe et al., 2015	Omer et al., 2016	ISRCTN34757603	positive	effectiveness

Table 2.9 lists some of the characteristics of the **OE** associated with the identified **PE**. Most of the studies were registered with a trial registry (87%) and had some form of blinding (63%). This indicated that the studies largely adhered to good trial practices. The studies mainly used the parallel **RCT** design (97%), had two trial arms (80%), adopted simple randomisation (50%) and were focused on effectiveness (73%). This indicated that the **OE**s study designs were fairly homogeneous. Lastly, about half of the studies reported a positive outcome (47%) suggesting that the researchers' motivation for conducting **PE** is irrespective of the main trial outcome. Table 2.10 summarises the information extracted from the **PE** studies. Most of the **PE** studies employed mixed methods (43%) or qualitative methods (43%) to conduct their evaluation. This suggests a strong methodological preference towards qualitative approaches in the conduct of **PE**. However, about a third (30%) of the studies had causal inference either as a stated or an implied aim. If we breakdown the methods used by whether or not causal inference was an aim (table 2.11), we could see that studies that employ solely qualitative methods do not consider causal inference as their aim and for studies that employ some form of quantitative methods, almost half of them do not have causal inference as an aim (8 yes vs. 9 no). We will revisit this in the discussion section.

Several kinds of processes were identified in this review. The names of the processes given in table 2.10 were either verbatim from the studies or matched with a closely named process. As such, some of the processes express related and overlapping concepts and I would briefly explain what each refers to. Ranking the processes by the number of studies evaluating it, 'subjective response to treatment (patients)' is the most commonly studied process (63%). This is an assessment of the patient's perception towards the treatment and whether or not there were particular aspects of the treatment that was pleasant or unpleasant. This is followed by 'adherence' (30%) and 'fidelity' (27%). Adherence refers to the patient's adherence to the treatment protocol while fidelity refers to the degree to which the treatment was implemented as planned. 'Subjective responses to treatment (Staff)' refers to the impression of the treatment by the personnel involved with delivering the treatment. 'Dose received' refers to the dosage of the treatment that the patient received and this is in contrast to 'dose delivered' which refers to the dosage that was delivered to the patient. There is a distinction because not all the dosage that was delivered would be received by the patients and this can happen for various reasons including non-adherence to the treatment protocol. Reach refers to the degree to which a treatment had been delivered to a target population. This is most commonly expressed as

Table 2.9: Characteristics of outcome evaluation

Characteristic	Count	%
Trial registration		
Yes	26	87%
No	4	13%
Presence of blinding		
Present	19	63%
Unstated / none	11	37%
RCT design		
Parallel	29	97%
Step-wedged	1	3%
Number of trial arms		
2 arms	24	80%
3 arms	6	20%
Type of randomisation		
Simple	15	50%
Cluster	12	40%
Stratified	3	10%
Aim of trial		
Efficacy	8	27%
Effectiveness	22	73%
Trial outcome		
Positive	14	47%
Negative	16	53%

Table 2.10: Characteristics of process evaluation

Characteristic	Count	%
Methods used in inquiry		
Quantitative	4	13%
Qualitative	13	43%
Mixed methods	13	43%
Causal inference as stated aim		
Yes	7	23%
No	23	77%
Causal inference as stated or implied aim		
Yes	9	30%
No	21	70%
Processes investigated		
Subjective responses to treatment (Patient)	19	63%
Adherence	9	30%
Fidelity	8	27%
Subjective responses to treatment (Staff)	7	23%
Dose received	4	13%
Reach	4	13%
Satisfaction	3	10%
Context	2	7%
Dose delivered	2	7%
Implementation	1	3%
Mechanisms of treatment	1	3%
Recruitment	1	3%
Results from **PE** linked to results from **OE**		
Yes	3	10%
No	27	90%

Table 2.11: Cross tabulation: Methods used x Causal inference as aim

	Causal inference as aim	
Methods used	Yes	No
Mixed methods	8	5
Qualitative	0	13
Quantitative	1	3
Total	9	21

a percentage of the target population who had received the treatment. Satisfaction refers to how well the treatment has met a patient's goal from their own perspective. Context refers to the environmental influences that dictates the conditions under which the treatment was provided. This is commonly studied to compare the effect of the same treatment under different conditions. An example would be the effect of a treatment when it is provided by a nurse versus a psychiatrist or when a treatment is delivered in a hospital setting versus a clinic in the community. Implementation refers to the processes involved in setting up the delivery of the treatment. Mechanisms of treatment refer to the processes that form a causal chain of events leading up to the production of the intended treatment effects. Recruitment refers to the processes related to the recruitment of patients during the trial.

Each of these processes occur during distinct stages of the **RCT**. Using these stages, a grouping of these processes was proposed:

1. Processes of trial conduct (recruitment, reach),
2. processes relating to context (context),
3. processes related to delivery of treatment (subject/staff responses to treatment, adherence, fidelity, dose received/delivered, satisfaction, context, implementation) and
4. mechanisms of action of the treatment (mechanisms of treatment).

This grouping serves to provide a structure to consider how processes from different parts of the **RCT** can affect the outcomes of interest and grouping helps to ensure that entire groups of processes would not be overlooked. This grouping is by no means a definitive way to group processes and it exists solely to aid and encourage consideration of the role of different kinds of

processes in an **RCT**. Lastly, 90% of the **PE** studies did not make any reference to the outcomes measured in the **OE**. This suggested that in the majority of the **PE** studies, the processes were evaluated independently from the outcomes.

2.4 Discussion

The primary aim was to understand what mental health researchers were using **PE** for and the methods associated with this use. From this review, we identified a broad range of processes that mental health researchers consider relevant. The methods associated with evaluating these processes were largely qualitative in nature even though in about a third of them, causal inference was an aim. These findings indicated that the interest in **PE** amongst mental health researchers could be driven by the unique nature of mental health treatments where processes such as the patient's 'perception of the treatment' can play a critical role in achieving beneficial outcomes as compared to evaluations of the pharmaceutical treatments where such perceptions of the treatment likely play a far lesser role. A grouping of these processes was also proposed to aid in thinking about the different kinds of processes that are present within a trial. The aim of this grouping was to encourage the consideration of the role of processes within each facet of the trial and its relation with the outcomes of interest. This role would in turn dictate the types of causal questions that could be posed in relation to the effect of the process on the outcome. The type of causal inference questions that could be posed falls into one of the three following types:

1. Mediation: Did the process mediate the effect of the treatment on the outcome?
    - Such research questions stipulate a pathway by which outcome improvements come about.
    - They seek to establish whether changes in outcomes are brought about by the treatment through an intermediate variable.
    - The aim of such questions is to ascertain how the observed changes in outcomes came about: directly or indirectly via intermediate variables.
2. Effect modification: Did the post-intervention process modify the effect of the treatment on the outcome?
    - Effect modification is similar to mediation in that it also seeks to determine if the changes observed in the outcome is due directly or indirectly to the treatment.

- Such research questions aim to establish whether the intervention is more beneficial for patients who experience the treatment in a certain way, e.g. taking part in a sufficient number of session (adherence) or having a better rapport with the therapist.
- Importantly such therapy experience variables are only observed by those who are offered the respective therapy. Such questions are again about establishing who can benefit from the intervention, and typically inform further complex intervention development.

3. Moderation: Did the process moderate the effect of the treatment on the outcome?
    - Moderation focused questions seek to establish whether the causal effect of the intervention varies amongst groups within the population (contextual variables).
    - These questions help to determine who is likely to benefit from the intervention and amongst those who benefit, which group of the population have the largest benefit.

Processes of trial conduct and context could pose either questions related to mediation or moderation. Processes related to delivery of treatment could pose questions related to effect modification and processes related to mechanisms of action of the treatment could pose questions related to mediation. In essence, this grouping of processes not only maps out the different processes conceptually, but also provides guidance on the type of causal questions that could be asked and by extension the methods that could be deployed to study these processes. All of these would come into play if causal inference was an aim of conducting **PE** but the findings of the review seem to suggest that causal inference was of little interest.

This was most evident in the heavy use of qualitative methods among the studies reviewed. Qualitative and quantitative methods differ starkly in their perspectives of reality where the former considers reality to be a social construct, shared by many individuals and the latter considers the existence of a single reality (Guba & Lincoln, 1981). This is an oversimplified characterisation of the perspective of reality of the two sets of methods. However, this was done to highlight how different each set of methods considers to be reality and by extension the means that each deploy to understand and find out about this reality.

Qualitative studies typically following along one of the five approaches (narrative, phenomenological, grounded theory, ethnographic and case study research) and each approach considers reality in different ways and holds different assumptions (Creswell, 2007). The findings from

qualitative studies can be considered as detailed accounts of this reality. For example, in the phenomenological approach, it considers reality to be informed by the collective experiences of many individuals as they go through certain phenomena or events. The aim of a study employing a phenomenological approach would then be to give a detailed account of this collective experience that is unique to this phenomenon. Such an account is a subjective evaluation of reality, a reality experienced by the individuals involved. These subjective assessments, in the context of **PE**, inform the various stakeholders of what it is like to going through specific processes and allows for a richer understanding of what happens at each part of the trial from different perspectives.

An important consequence from this richer understanding is that it aids in forming new hypotheses about causal relationships amongst processes and between processes and outcomes. These new hypotheses could indicate potential ways in which the treatment could be modified to bring about stronger treatment benefits. The shortcoming of a purely qualitative approach is that there is no objective measure of the degree in which the new hypotheses are to be trusted. The is where quantitative methods and specifically, methods of causal inference can aid in providing an objective assessment of this hypothesised causal relationship.

Using a drug trial as an example, a qualitative finding could indicate that some patients did not adhere to the medication regime because of the bitter taste of the drug. A hypothesis could be that the bitter taste caused some patients to miss their doses and the lapse in dosing caused a drop in the treatment effectiveness. If such a drop in effectiveness was deemed important to remediate and a change in the composition of the drug was trivial, the drug could be changed to taste less bitter without any further action. However, if this change was not trivial, then the prudent thing to do would be to generate evidence in support of this hypothesis before any changes were made. Such would also be the case if there was a complex web of causation where the change in a single variable could have unexpected or unpredictable effects as in the case of complex interventions. The testing of this hypothesis could either be done using existing data if the necessary data was available or in a new trial designed specifically to evaluate the impact of the taste of the drug on its effectiveness.

Having stated the role of qualitative and quantitative methods in the context of an **RCT**, I would address the sparing use of quantitative methods in **PE** next. This sparing use could be explained

either by a lack of need for questions of causality in **PE** or that there is such a need but barriers exist in conducting such an analysis. Addressing the former, it seems unlikely since the aim of an **RCT** was to identify better options for treatment than existing ones. Furthermore, focusing on the processes that were of interest in the studies reviewed, many of them such as 'adherence' and 'subject's perception towards treatment', had an unambiguous potential causal relationship with the outcome. We can therefore rule out a lack of need for causal inference in **PE**.

Having ruled out the lack of motivation for the use of quantitative methods, the existence of barriers becomes the more likely explanation for the lack of use of quantitative methods for causal inference in **PE**. This is supported by the finding that only 10% of the studies reviewed sought to infer some relationship between the processes studied and the outcomes of interest. Given that many of the processes had potential causal relationships with the outcomes as indicated previously, the lack of studies seeking to test this hypothetical relationship suggests that this was a task with significant difficulties.

If we also consider the type of causal questions that could be posed in a **PE** and the methods that could be employed to test the hypotheses posed by those questions, it becomes clear that for a subset of these questions, such as some questions of mediated causality, accessible tools are not available for use. Accessible in this context refers to tools that do not have prerequisite training or in-depth knowledge about statistical methods and programming in order to use. This lack of accessible tools is another indication supporting the notion that the observed sparing use of quantitative methods is secondary to the barriers encountered. The lack of accessible tools apply to a subset of questions posed in **PE** and the next chapter would focus on questions of mediated causality in **PE**.

Having reasoned about the existence of barriers in the conduct causal inference within **PE**,the next few chapters would focus on elaborating and overcoming some of these barriers.

2.5 Strengths and limitations

A limitation of this study was the choice to focus only on studies related to mental health. The main motivation behind this choice was to be able to identify and develop methods for **PE**

within the mental health domain. By doing so however, it also implied that the conclusions drawn from this review had limited generalisability beyond mental health studies. This brings up the question of whether significant changes to the findings of this review would be expected if we expanded the scope of this review to include all studies that utilise **PE** and more specifically whether the methodological gaps uncovered would have already been plugged.

There are two indications that this is not the case. The first, as mentioned previously, is a lack of easily accessible tools to conduct causal inference for common scenarios found within **PE**. The second indication comes from the manual screening process where each study was assessed to determine if it fulfilled the inclusion criteria. Since there were no keywords within the search strategy to identify mental health related studies, this meant that the studies that were retrieved from the initial search had to be screened to confirm that it was a mental health related study. From the 930 studies that passed through screening, less than 10 studies fulfilled all the inclusion criteria except for being a mental health related study. Amongst these studies, there was no indication that the methodological gap in causal inference had been filled. Therefore, in a bid to maintain the focus of this thesis and to allow the review to contextualise **PE** as it applied to a mental health researcher, these studies were not included in this review. Lastly, given the small number of studies that were not mental health related, there is high degree of confidence that the conclusions drawn from this review will remain largely unchanged with the inclusion of these studies.

The strength of this study lies in its transparency. It is possible to reproduce in entirety the results of this review using the reviewer's manual found in the appendix where detailed documentation of each step of the search was documented and parts of the search was additionally verified by a second reviewer. Another strength of this study lies in its novelty. This study taps on many of the same steps as a traditional systematic review but for the purpose of identifying a methodological gap in the literature. Using evidence from existing literature to identify and motivate methodological developments ensures that methods arising from these efforts are better positioned to fulfil a genuine need within the research community.

2.6 Conclusion

In conclusion, this review achieved its two aims of identifying the purposes of **PE** as well as the methods used to accomplish the purpose. The results from the review indicated that processes were studies with the aim of improving treatment outcomes. The sparing use of causal inference methods in **PE** is a sign that there are methodological barriers. A consideration of the questions posed in **PE** provided hints that these methodological barriers exist since the tools required to answer the questions are not easily accessible. The rest of this thesis would focus on overcoming some of these barriers in a bid to encourage greater scrutiny of processes as a means to a deeper understanding of why, how and for whom does a treatment work. Lastly, the review also propose a grouping of processes with the aim of ensuring that processes from different parts of the **RCT** were given due consideration as to their role in affecting treatment outcomes.

Chapter 3

A framework for causal mediation analysis

3.1 Introduction

The previous chapter concluded that many research questions posed in process evaluation (**PE**) are of a causal nature. These questions arise from a need to understand the context and pathways through which interventions cause improvements in outcomes. They can be broadly categorised into three types of causal questions discussed previously: *mediation*, *moderation* and *effect modification* questions. Out of the three types of questions, mediation questions are perhaps most pertinent to **PE** since processes commonly serve as intermediaries between the interventions and outcomes. This idea of thinking of processes as mediators is not new and had been highlighted by Judd and Kenny (1981), Judd (1987), and also in the seminal paper on mediation by Baron and Kenny (1986). The first two papers also referred to the use of mediation as a way to conduct *process analysis*, the name used to refer to **PE** in an earlier time.

Given these beginnings of **PE**, it comes as an unexpected finding that qualitative methods were heavily used in the evaluation of processes and not quantitative methods since quantitative methods are well positioned to answer questions posed by **PE**. Consequently this also implied that when qualitative methods were used, researchers were able to only generate new hypotheses from the results of the studies but were unable to confirm them. Confirmation of these

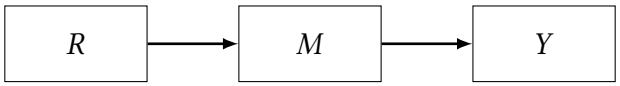
hypothesises required the use of quantitative methods and the challenges of doing so were briefly discussed in the previous chapter. What appeared to be most likely is that causal inference was of interest but owing to methodological difficulties of causal inference, qualitative methods were used instead. This however limited the use of **PE** to provide insight into how the processes work to cause an effect of interest. This is particularly important for complex interventions. Complex interventions, by their nature, consist of each multiple components, each with different effects and the interplay of these effects sometimes have unexpected or unknown effects on the outcomes of interest (Moore et al., 2015b). This interplay makes it difficult to discern how exactly the outcome of interest came about and to which component of the intervention can one attribute the outcome of interest to. This applies regardless of whether or not the trial shows a desirable outcome or not. Should the trial show that the treatment is ineffective, its also of interest to understand why this came about. In this respect, **PE** can play an important role to understand with precision the factors contributing to the success or failure of complex interventions.

Given the promise of **PE** in aiding an understanding of causal relationships in complex interventions coupled with the potential of conducting such analyses using existing data, this thesis aims to develop methods to bridge this methodological gap with a focus on the assessment of causal mediation. Trials of mental health interventions will be used to guide the methods development to ensure that methods developed had real world applicability in an existing, active domain.

Up till this point, mediation has been discussed in the context of causal inference as this is the emphasis of this thesis. There is another concept of mediation termed statistical mediation (MacKinnon, 2008). Both statistical and causal mediation are aimed at establishing causality. The difference however is that for causal mediation, the aim is explicit and the mediational analysis would be conducted within a causal framework such as the potential outcome (**PO**) framework. Statistical mediation on the other hand have an implied causal interpretation of mediation and focuses on associations and correlations. These would be considered to be weak evidence for causality using the ladders of causation developed by Pearl (2019). Causal mediation would be higher up the ladder of causation and represents stronger evidence of causality which would be discussed in greater detail in this chapter. Both statistical and causal mediation however share similar statistical methods in estimating the degree to which an effect of interest is mediated and these include the generalised linear model (**GLM**) family of models to estimate the association

between the treatment under study R and the target intermediate outcome/putative mediator variable, M, as well as between M and the distal outcome variable, Y (fig. 3.1). However, in order to infer any causality, a quantification of the causal effect, the causal estimand, needs to first be defined. The causal estimand adopted in this thesis is a population parameter which we wish to estimate from a sample of subjects that we can observe. Estimation of the causal estimand then requires the selection of a statistical approach and the different approaches may hold different assumptions about the underlying attributes of the population. Focusing on the

Figure 3.1: Simple mediation diagram



difference between statistical and causal mediation, causal mediation has substantially more considerations since it goes beyond making associations to making causal inferences. These additional considerations required to conduct causal inference would be approached using principled methods by clearly defining the causal estimands and any assumptions made in the process of estimating the causal estimands (Daniel & De Stavola, 2019; De Stavola & Daniel, 2012; Pearl, 1995). Causal mediation analysis refers to such principled approaches that can produce valid (unbiased) estimates of mediation estimands (T. J. VanderWeele, 2015).

This chapter is broken up into several sections. The next section (section 3.2) introduces the framework for causal mediation analysis used to define mediated effects that might be of interest in the context of **PE**. Following on, the next section (section 3.3) makes a detour to provide some background information on a randomised controlled trial (**RCT**), the Carers' Assessment, Skills and Information Sharing (**CASIS**) trial, which serves as a motivating use case to illustrate how the various concepts apply to an **RCT**. **CASIS** is also used in the next chapter in the development of the estimators. Discussions are limited to the case of two-arm **RCT**s as this is the primary focus of the thesis.

The following section (section 3.4) formally defines total, mediated and non-mediated causal effects of interest in a target population. The definitions are discussed alongside issues related to identification of the effects. Identification here refers to a specific concept in statistics where a population quantity is identified if it can be expressed as a function of moments (means and covariances) of observable variables. The quantities of interest in the context of **PE** are the

total or mediated causal effects in the population. Identifiability thus refers to the conditions under which the variables need to be observed under in order to be able to compose the function relating the population quantity of interest with the observed variables. These conditions and the observed variables depend upon the study design. If the causal effect cannot be observed under any conditions, it implies that the causal effect cannot be estimated at all and such a population quantity is referred to as a non-identified causal effect. Having an identified causal effect thus implies that under appropriate conditions, the causal effect can be estimated.

After defining the causal effects of interest and addressing identification issues, I will next discuss how the defined causal effects can be estimated from sample data using fully parametric methods. Parametric models refer to statistical models which parametrise or quantify all distributions of random variables and relationships between variables. Only inferential methods relying on fully parametric models are considered because the parametric assumptions inherent in such models aid in identifying causal effects and provide for more efficient estimation approaches as compared to non-parametric approaches. The validity of the results from parametric approaches however rests upon how likely the parametric assumptions are violated in a given study.

This chapter ends with a summary of how mediated causal effects are defined, identified and estimated. This lays the groundwork for developing methods that address unmet methodological needs which is the focus of the following chapter.

3.2 Potential outcomes framework

The framework that is adopted to define, identify and estimate causal effects is the **PO** framework. In this section, I briefly introduce main elements of the **PO** framework. The **PO** framework has been used in recent times to define what constitutes a causal effect (Pearl, 2009; Rothman et al., 2008) and different names have been used to refer to it: the counterfactual framework (Lewis, 2001), the Neyman-Rubin causal model, the Rubin Causal Model (**RCM**) (Holland, 1986) and the structural causal model (**SCM**) (Pearl, 2009). These different names all refer to the same concept of using potential outcomes to define what a causal effect is. It should be noted however the conception of using **PO**s behind each of these frameworks are not exactly the same and the definition adopted in this thesis closely aligns with the framework by Pearl (2009).

The word 'potential' in 'potential outcome' indicates an outcome that can occur but has yet to occur. In the context of a two-armed **RCT**, the **PO** refer to outcomes attributed to each of the treatment arms. Such a trial typically has an experimental and a treatment as usual (**TAU**) arm where the latter is used as a control. For simplicity, for this definition, I am assuming that the treatment offered or assigned to the subject is the one that he or she receives. This may not be true under certain circumstances, for instance, if the subject refuses the treatment. The reason this is important is because in a typical trial, there is a distinction between the causal effect on treatment offered (or more commonly known as intention-to-treat (**ITT**) analyses) and the causal effect on treatment received (or more commonly known as the complier average causal effect (**CACE**)). The investigator may be interested in one or the other and therefore defines the causal effect accordingly. For now, I will assume that the causal effect of interest is the **ITT** effect and the causal effect can be defined for everyone in the trial regardless of what arm they are in. The **PO**s are defined as follows:

1. the outcome had the subject been offered the experimental treatment and
2. the outcome had the subject been offered the control treatment.

Using these two **PO**s, we can then construct a causal effect. As an example, let the outcome be a measure of severity of a disease with a higher score indicating increased severity. The **PO**s in this case mean what the severity of disease would have been should the subject be offered the different forms of treatment. For a continuous measure such as severity, a useful way to define an individual's causal effect of treatment is to compare the potential outcomes under experimental and control treatments. This can be done by simply taking the difference between the **PO** for the experimental and control treatment. If this difference is positive, it indicates that the experimental treatment causes the disease to be more severe and if it is negative, it indicates that the experimental treatment is reducing the severity of the disease. If there are no differences in the two **PO**s, it indicates that the experimental and control treatments are on par in terms of the **PO**s.

More formally, Y refers to the outcome and R refers to the offered treatment where,

$$R = \begin{cases} 1, & \text{if offered experimental treatment} \\ 0, & \text{if offered control treatment} \end{cases}$$

When an individual is being referred to, a subscript would follow like Y_i referring to the outcome

Y of subject i. Using these notations, the two **PO**s for subject i can then be notated as $Y_i(R = 1)$ and $Y_i(R = 0)$ representing the **PO** of being offered the experimental and control treatment respectively. These two notations are abbreviated to $Y_i(1)$ and $Y_i(0)$ in this thesis. These two **PO**s are mapped onto the observed outcome Y_i as follows:

$$Y_i = \begin{cases} Y_i(1) & \text{, if } R = 1 \\ Y_i(0) & \text{, if } R = 0 \end{cases}$$

The causal effect for subject i on treatment offer, also known as the total effects $(\mathbf{TE})_i$, is thus,

$$\begin{aligned} \mathbf{TE}_i &\triangleq Y_i(R = 1) - Y_i(R = 0) \\ &\triangleq Y_i(1) - Y_i(0) \end{aligned}$$

Having presented a brief overview of the **PO** framework, I will next provide some background information on the **CASIS** trial which will serve to motivate the concepts presented in this thesis.

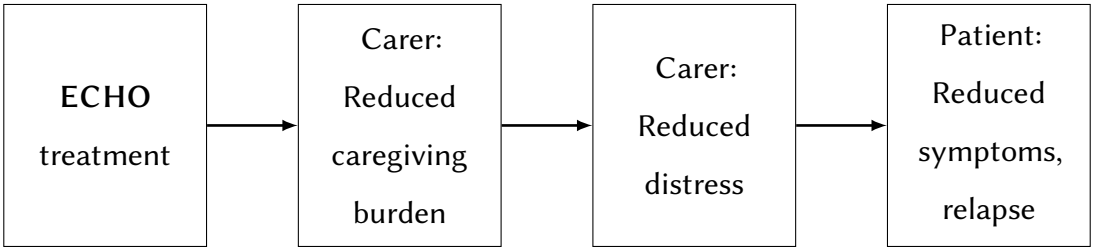
3.3 The CASIS trial

This section presents the background to the **CASIS** trial (Goddard et al., 2013; Hibbs et al., 2015; Magill et al., 2016), which is used to illustrate the applicability of various concepts discussed throughout the thesis to an **RCT**. **CASIS** was a trial of an intervention, Experienced Carers Helping Others (**ECHO**), for carers of patients with anorexia nervosa (**AN**) which is a form of eating disorder characterised by a persistent restriction of energy intake, fear of gaining weight and an altered self-perception of one's own weight. **AN**, in severe cases, can become life threatening (American Psychiatric Association, 2013). The **ECHO** is a programme aimed to improve a carer's ability to care for a person with **AN** and to reduce the severity of the symptoms of **AN** exhibited by the patient. For the **CASIS** trial, patients who have been diagnosed with moderate to severe **AN** were recruited.

The **ECHO** is a theory driven self-help intervention designed to reduce adverse interactions between carer and patient as well as to support the well-being of the carer. The theory postulates that by achieving these two goals, the carer will have decreased levels of distress, leading to decreased occurrences of relapse and fewer **AN** symptoms in the patient. A simplified diagram

of the theoretical causal relationships is shown in fig. 3.2. The **ECHO** programme consisted of

Figure 3.2: Simplified causal diagram of **CASIS** (Goddard et al., 2013)



a book and five DVDs providing communication skills training targeted at supporting persons with **AN**. The programme was supplemented with telephone coaching provided by experienced caregivers or coaches who received prior training in the same communication strategies taught in the book and DVDs. These caregivers and coaches were assessed by the researchers to have reached a pre-requisite level of competency before they started providing coaching sessions. The treatment took place over a period between three to six months and the carer-patient dyad/triad were followed up for a year.

All outcome measures were assessed on admission, discharge to and from the hospital, 6, 12, and 24 months post-discharge with the exception of body mass index (**BMI**) of the patient which was measured monthly. The trial had separate aims for the carer and the patient. For the carer, the primary aim was to decrease the level of distress and this was measured using the Depression, Anxiety and Stress Scale Short Form (**DASS-21**) (Henry & Crawford, 2005; Lovibond et al., 1995) with the primary **DASS-21** outcome defined at 12 months post-discharge. The primary aim for the patient was to increase the time to relapse (second co-primary trial outcome) and this was monitored by **BMI**. If the patient was at normal weight at the time of discharge, a relapse was defined as dropping below 17.5kg/m^2. If the patient was discharged at below normal weight, a drop of 1kg/m^2 indicates a relapse.

The secondary outcomes for the caregiver were as follows:

1. the Accommodating and Enabling Scale for Eating Disorders (**AESED**) (Sepulveda et al., 2009) as a measure of accommodating behaviours which refer to the accommodation of behaviours associated with **AN** exhibited by the patients,
2. the Eating Disorders Symptom Impact Scale (**EDSIS**) (Sepulveda et al., 2008) as a measure of caregiving burden,

3. the Family Questionnaire (**FQ**) (Wiedemann et al., 2002) as a measure of expressed emotion which refers to certain attitude and behavioural patterns shown towards the patients,
4. the World Health Organization Quality of Life (**WHOQOL**) (World Health Organisation, 1996) as a measure of quality of life (**QOL**) and
5. the Client Service Receipt Inventory (**CSRI**) (Beecham & Knapp, 2001) as a measure of resource use.

The secondary outcomes for the patient were as follows:

1. the **DASS-21** (Henry & Crawford, 2005; Lovibond et al., 1995) as a measure of distress,
2. the Eating Disorder Questionnaire (**EDE-Q**) (Bryant-Waugh et al., 1996; Watkins et al., 2005) as a measure of symptoms related to **AN**,
3. the **WHOQOL** (World Health Organisation, 1996) as a measure of **QOL** and
4. the **CSRI** (Beecham & Knapp, 2001) as a measure of resource use.

A detailed timeline of the assessments will be presented in the following chapter where the developed methods are applied to the data from the **CASIS** trial. The trial adopted an **RCT** study design. A total of 178 patients and 268 carers were recruited into the study. The recruitment took place upon the patient being admitted to the hospital for an **AN** related event and the randomisation of the subjects took place right after recruitment. The randomisation of the treatment was at the level of the patient-carer(s) dyad or triad. 92 patients and 134 carers were randomly offered to receive **TAU** only and 86 patients and 134 carers were offered to receive both **TAU** and **ECHO**.

The original analysis in the study catered for more than one caregiver but for the purposes of illustrating the concepts presented in this thesis, I will simplify this and limit the analysis to only the primary caregiver. The primary caregiver is a distinction made within the study to identify the main person that was in charge of the care of the patient. The study also conducted a qualitative study focused on the experience of caregivers and patients undergoing **ECHO** closely aligned with the process 'subjective perspective of treatment (patients)' identified in the previous chapter (Macdonald et al., 2014). The main trial indicated that the outcomes for both the carers and patients were marginally and non-significantly better in the **ECHO** arm of the trial. The qualitative study indicated that both the carers and patients found the support afforded by **ECHO** helpful and both also had better insights into their communication with

each other. With this introduction to the **CASIS** trial, I now proceed with formally defining causal treatment effects.

3.4 Definition of mediated causal effects

Prior to the introduction of **CASIS**, we left off from a discussion of the **PO** framework and how it is used to define causal effects. In this section, I pick up from where we left off and first expand on the use of the **PO** to introduce two different total causal effect size measures for continuous and binary outcome variables respectively: the causal mean difference and the causal odds ratio (**OR**). Following on from this, I will introduce a formal definition of the mediated causal effect with a single mediator, first with continuous causal effects defined using differences and then causal **OR** for binary outcomes. Finally, I will expand the definition of a mediated causal effect to allow for two sequential mediators. Identification issues with the causal effects are presented alongside their definitions.

3.4.1 Total causal mean differences and OR

In the section on the **PO** framework, the **TE** for a single individual was introduced. A brief recap, consider a two-arm **RCT** with Y being the outcome of interest and R being the treatment offered where $r = 0$ when the control treatment had been offered and $r = 1$ when the experimental treatment had been offered. The **TE** for an individual, $\mathbf{TE}_i$ is thus defined as:

$$\mathbf{TE}_i \triangleq Y_i(1) - Y_i(0)$$

From this definition, we can see that in order to estimate an individual causal effect, we will need the outcomes of the same individual after being offered the experimental and control treatments of the trial concurrently. This is not possible since there is only one individual and he or she can only be offered a single treatment at any given time. A scenario that comes close to being able to measure the outcomes of both arms of the trial for each individual concurrently is the cross-over trial where each subject receives both the treatments one after another. In most cases, this serves as a good approximation of concurrently measuring both **PO**s. The key difference is that the individual would not be physiologically or psychologically identical at the start of

both treatments in the cross-over trial. The cross-over trial will not be discussed further as it is a trial design used only in select circumstances. The focus will be placed on a standard two-arm parallel group **RCT**. In such a typical **RCT** setup, the inability to measure both **PO**s at once indicates that the individual causal effect cannot be estimated and is therefore unidentified.

A way around this inability to estimate an individual treatment effect in a standard **RCT** is to consider causal treatment effects treatment effects at the population level. We define the population total causal effect of treatment offer **TE** using the population causal effect as follows:

$$\mathbf{TE} \triangleq \mathrm{E}[\mathbf{TE}_i] \tag{3.4.1}$$

$$= \mathrm{E}[Y_i(1) - Y_i(0)] \tag{3.4.2}$$

It should be noted that due to the linearity of expectation, eq. (3.4.2) is equivalent to the difference between the means of **PO**.

$$\mathbf{TE} \triangleq \mathrm{E}[Y_i(1)] - \mathrm{E}[Y_i(0)] \tag{3.4.3}$$

Therefore, taking the expectation of differences between **PO** of each individual eq. (3.4.2) and taking the differences between the expectations of each **PO** eq. (3.4.3) are equally valid and equivalent expressions of the **TE**. The next question of interest would be whether the **TE** is identified? When considering identification, we have to take into account the study design and in this case, a parallel group **RCT**. In such a design, subjects are randomly offered either the control or the experimental treatment. The randomisation of the treatment offered ensures that the characteristics of both groups are similar at baseline and because their characteristics are similar at baseline, the outcomes associated with the control and experimental treatment will stay the same if the two groups of subjects exchanged the treatment they were offered, i.e. the control treatment group receiving the experimental treatment and the experimental treatment group receiving the control treatment. This unique property of being able to *exchange* the offered treatment and still have consistent observed outcome is known as *exchangeability* or *ignorable treatment assignment condition* (Hernan & Robins, 2020). This property is ensured in trials due to randomisation. What exchangeability allows us to do is to be able to use simple means and differences in accordance with eq. (3.4.2) or eq. (3.4.3) to obtain valid estimates of **TE**.

However, being able to estimate the **PO**s alone does not make the **TE** an identified causal effect under the standard **RCT** design. Exchangeability is the first of the three conditions of

identification that must be fulfilled before the **TE** can be considered identified. I will take a short detour to discuss the other two conditions before continuing with the definitions of causal effects.

The second identification condition, known as *positivity* (Hernan & Robins, 2020), requires that for each individual, there is a non-zero chance of being offered each of the treatment under investigation within the trial. In the context of an **RCT**, positivity holds by design as each of the treatment that is randomly offered has a predefined non-zero probability of being assigned to each individual.

The last condition for identification is *consistency* (Hernan & Robins, 2020). Consistency requires that the observed outcome for each individual to be equivalent to the outcome had the individual received the offered treatment and is implied by equation eq. (3.4.1). On the surface, this condition appears to be a given but it can be violated if the offered treatment is ill defined, allowing for wide variations in administration of the treatment. In the case of **CASIS**, if the carers each received a different version of the self-help guide or if the coaches provided support to each of them in very different manners, it then becomes unclear what causal treatment effect is targeted. This is more likely to be violated when the interventions are complex or consist of multiple parts and less likely when the intervention is straightforward for example in a typical drug trial. In summary, what consistency requires is for different treatments within the trial to be well-defined and adhered to.

Moving back to the definition of **TE**, we have so far defined **TE** using differences between two **PO**s. Such a definition is meaningful when the outcome is continuous. When the outcome is binary, the group level **PO**s becomes a proportion of the subjects who had an outcome of 1 within the group. A direct consequence of this is that the **TE** becomes the difference in probability of having an outcome of 1 between the experimental and control treatment.

$$\begin{aligned}\mathbf{TE} &\triangleq \mathrm{E}[Y(1) - Y(0)] \\ &\triangleq \mathrm{E}[Y(1)] - \mathrm{E}[Y(0)] \\ &\triangleq \Pr(Y(1) = 1) - \Pr(Y(0) = 1)\end{aligned}$$

Since each of the **PO**s is a probability, an alternative way of defining a **TE** is by using an alternative effect size measure, the causal **OR**. The **OR** is a popular way of expressing the odds that one would recover when given an experimental as compared to a control treatment. Its

popularity likely stems from the widespread use of logistic regression to analyse case control studies. In order to define the causal **OR**, let's begin by first defining what odds are. Odds is a ratio between the probability of having an outcome of 1 under a specific treatment and the probability of having an outcome of 0 under the same treatment. Therefore, the odds of $\mathrm{E}[Y(1)]$ are:

$$\frac{\Pr(Y(1)=1)}{\Pr(Y(1)=0)}$$

The same applies for the second **PO** in the definition, $\Pr(Y(0)=1)$:

$$\frac{\Pr(Y(0)=1)}{\Pr(Y(0)=0)}$$

With the two odds defined, we can now define the total causal **OR**. The **OR**, as the name implies, is a ratio of odds. Using the two population odds that had just been defined, the **OR**, known as the average treatment odds ratio (**TOR**) is:

$$\mathbf{TOR} \triangleq \frac{\Pr(Y(1)=1)}{\Pr(Y(1)=0)} \Big/ \frac{\Pr(Y(0)=1)}{\Pr(Y(0)=0)}$$

The **TOR** is the equivalent to the **TE** but for binary outcomes and serves to be a measure of the total causal effect of the treatment on the outcome in terms of **OR**. The **TOR**, given that it is a ratio of odds, indicates that if it is 1, the two odds in the ratio are the same. If it is below 1, the odds from the denominator which represents the odds for the control treatment is smaller than that of the numerator which represents the odds for the experimental treatment. The vice versa is true as well. In a standard **RCT**, the **TOR** is identified provided all of exchangeability, positivity and consistency hold for the same reasons as explained above for the **TE** effect size for continuous outcomes.

Collapsibility of an effect size measure refers to whether or not the constant conditional effect across subpopulation is the same as the marginal (whole population) effect. One important difference between the **TE** and the **TOR** is the property of collapsibility. Importantly, while the **TE** is collapsible, the **TOR** is not (Berzuini et al., 2012). In what follows, I will discuss relevant aspects of collapsibility as it applies to identifying a strategy for its estimation.

Consider the **TE** where the causal effect is defined in terms of differences between two expected **PO**s. The **TE** is the treatment effect of the population, formally known as the marginal treatment effect. We can also define a total causal effect in a subpopulation indexed by $X = x$ as follows,

$$\mathbf{TE}_x \triangleq \mathrm{E}[Y_i(1)|X=x] - \mathrm{E}[Y_i(0)|X=x]$$

This is referred to as the conditional causal treatment effect as it is conditional on variable X taking the value x where X is an indicator of the subpopulation. Using gender as an example, a conditional treatment effect can refer to the **TE** of males when X = males. Likewise, we can also have a conditional treatment effect that refers to the **TE** of females. If we know the proportion of males and females in the population as well as their respective **TE**, we would be able to work out the marginal treatment effect from these two pieces of information alone as shown in eq. (3.4.4).

$$\begin{aligned}\mathbf{TE} \; = (&\text{Proportion of males} \cdot \mathbf{TE}_{\text{males}}) + \\ &(\text{Proportion of females} \cdot \mathbf{TE}_{\text{females}})\end{aligned} \tag{3.4.4}$$

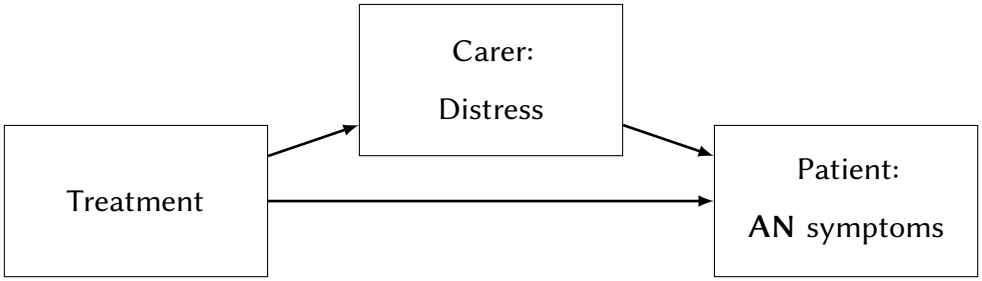
In the special case where $\mathbf{TE}_{\text{male}}$ is the same as $\mathbf{TE}_{\text{female}}$, gender is thus not an effect modifier and eq. (3.4.4) indicates that **TE** is equal to both $\mathbf{TE}_{\text{male}}$ and $\mathbf{TE}_{\text{female}}$.

An important difference between the **TE** and the **TOR** is that **OR** as effect size measures are in general, not collapsible. Even if the conditional **TOR** was constant across all subpopulation defined by $X = x$, it is still possible that $\mathbf{TOR}_x \neq \mathbf{TOR}$. It suffices for now to highlight that this is an important difference that will later pose a problem in estimation of the **TOR** and more will be elaborated in the next chapter. Collapsability is an important concept within causal inference with serious implications on interpretability of causal effects and a detailed discussion is well beyond the scope of this thesis. Such a discussion however can be found in Greenland and Pearl (2011) and Pearl (2009) and Hernán et al. (2011). Having defined the **TE** and **TOR**, we shall next define mediated causal effects for a single mediator.

3.4.2 Mediated causal effects for a single mediator

A mediated causal effect can be thought of as the effect due to a chain of events, and intuitively, this is akin to the dominoes effect where each domino falls as a result of the preceding domino falling on it. Using **CASIS** as an example, a mediated causal effect could be defined as the effect of the **ECHO** treatment in reducing symptoms of **AN** in the patient as a result of a reduction in distress of the caregiver by 12 months after discharge. The relationships of the three variables are shown in fig. 3.3. Each of the arrows represent a causal effect and in this case, since there are multiple arrows pointing towards the outcome 'Patient: **AN** Symptoms', it implies that there are

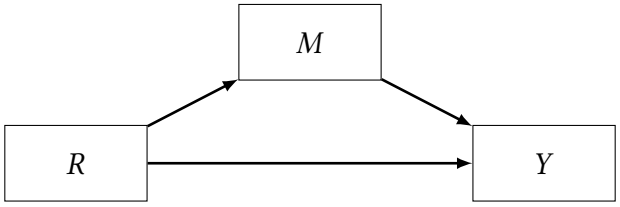
Figure 3.3: Mediated causal effects in **CASIS**, single mediator



multiple causes of this outcome. Tracing where these causes come from starting from the top lead us to 'Carer: Distress'. The 'Carer: Distress → Patient: **AN** symptoms' arrow represents the causal effect of carer's distress on the patient's **AN** symptoms. Since there is also an arrow pointing at 'Carer: Distress', this implies that there is a cause of the carer's level of distress. Tracing the arrows backwards, we see that this cause is the treatment. The 'Treatment → Carer: Distress' arrow represents the effect of the treatment on the levels of distress of the carer. We next turn to the bottom arrow pointing at the outcome 'Patient: **AN** Symptoms'. Tracing the origin of this arrow leads us again back to the treatment indicating that the treatment is a cause of the observed symptoms of the patient. The 'Treatment → Patient: **AN** Symptoms' arrow therefore represents the effect of the treatment on the symptoms of **AN** exhibited by the patient. Put together, we can see that the treatment can affect the outcome through different pathways. In the diagram, there is one direct and one indirect pathway from the treatment to the outcome. The indirect pathway is what is known as a mediated effect as it is mediated by an intermediary (Pearl, 2009). Since the **TE** of the treatment on the outcome remains the same regardless of how many mediators there are, a mediation model can be seen as a model that partitions or attributes this **TE** to each of the direct and indirect pathways. While it is self-evident, it is still worthy to highlight that a distinction between direct and indirect pathways is only relevant when both are present. In the absence of indirect pathways, there will be no such distinction and only the **TE** as defined earlier exists.

Next, to more easily refer to the each of the variables in a mediation model, I will use a mediation model as shown in fig. 3.4. R refers to the treatment, M refers to the mediator and Y refers to the outcome of interest. Figure 3.4 is simply a general version of fig. 3.3. For fig. 3.4, R represents the treatment, M represents the mediator and Y represents the outcome of interest. The arrows, as before, represent causal effects. The direct effect (**DE**) is represented by the $R \rightarrow Y$ path while the indirect effect (**IE**) is represented by the $R \rightarrow M \rightarrow Y$ path. The next issue to be addressed

Figure 3.4: Mediation diagram with 1 mediator



is how the **DE** and **IE** are to be defined using **PO**s. From fig. 3.4, we can see that there are two arrows pointing away from R indicating a causal effect of R on M and Y respectively. We also know that R can assume two values, 0 and 1. Starting with M, the two values of R indicates that there would be two M **PO**s, $M(0)$ and $M(1)$. In words, these are the values M would have had if R took the value of $r = 0$ and $r = 1$ respectively. Moving on to Y, we can see that there are two arrows pointing towards Y from M and R. Since there are two **PO**s of M and two possible values of R, the total combinations of M **PO**s with values of R would be 4, indicating that there would be 4 Y **PO**s. They are,

$$Y(R = 0, M = M(0)) = Y(0, M(0)) = Y(0)$$
$$Y(R = 1, M = M(1)) = Y(1, M(1)) = Y(1)$$
$$Y(R = 0, M = M(1)) = Y(0, M(1))$$
$$Y(R = 1, M = M(0)) = Y(1, M(0))$$

In words, $Y(0, M(0))$ would be the value that Y would have had if R took the value of $r = 0$ and the value of M is set at the value of M if R took the value of $r = 0$. Since both the r values in this *nested PO* are the same, this is a **PO** that can be observed in an **RCT** and it is the same as $Y(0)$. The nested **PO** $Y(1, M(1))$ is interpreted similarly and is the same as the **PO** $Y(1)$. Also, since $Y(0, M(0)) = Y(0)$ and $Y(1, M(1)) = Y(1)$, this also implies that the **TE** could be defined using these two **PO**s.

$$\mathbf{TE} \triangleq \mathrm{E}[Y(1)] - \mathrm{E}[Y(0)] \tag{3.4.5}$$
$$= \mathrm{E}[Y(1, M(1)) - Y(0, M(0))] \tag{3.4.6}$$

$Y(0, M(1))$ is the value that Y would have had if R took the value of $r = 0$ and the value of M is set at the value of M if R took the value of $r = 1$. In this case, the two r-values are different and this nested **PO** is a **PO** that is not possible to observe since we cannot subject M and Y to different levels of R at the same time. Such a **PO** is known as a *cross-world PO* and is

hypothetical in nature. Both $Y(1, M(0))$ and $Y(0, M(1))$ are *cross-world* **PO**s except for different combinations of r-values (Lewis, 2001; T. VanderWeele & Vansteelandt, 2014).

With the 4 **PO**s defined, we can now move on to define **IE** and **DE**. Previously, when defining **TE**, there were only 2 **PO**s, one each for the experimental and control treatment, therefore a choice need not be made over which **PO** to use to define the **TE**. However, in the case of the **DE** and **IE**, there are 4 **PO**s and only 2 are needed for each definition of a causal effect.

Starting from the **DE**, the first consideration is how should a **DE** be interpreted? A **DE** is an effect that does not pass through a mediator. Using this intuition, it follows that the effect should be defined using **PO**s where the value of the mediator if held at the same level of R in the two nested **PO**s contrasted. This leaves us with two possible ways to define the **DE**. These two ways were given special names, the *pure natural direct effect (**PNDE**)* and the *total natural direct effect (**TNDE**)* (T. VanderWeele & Vansteelandt, 2014).

$$\textbf{PNDE} \triangleq \mathrm{E}[Y(1, M(0))] - \mathrm{E}[Y(0, M(0))]$$

$$\textbf{TNDE} \triangleq \mathrm{E}[Y(1, M(1))] - \mathrm{E}[Y(0, M(1))]$$

The 'natural' in **PNDE** and **TNDE** refer to the effect of exposure to naturally occurring levels of the treatment on the mediator. In other words, it refers to the effects under the context of the value M set to the value it would have been under $r = 1$ or $r = 0$. This is contrasted with another form of effects known as the controlled effects where the level of M is held at a fixed, predetermined value for all subjects. The controlled **DE** has been widely used in policy analysis where the effect of an intervention in a hypothetical population with M set to a fixed value (i.e. the value expected due to a policy) is of interest (Pearl, 2009; Pearl et al., 2016). For example, if we had an intervention that improved the symptoms of an eating disorder via a reduction in anxiety, controlled **DE** could be used if we are interested in the direct effect of the intervention on the symptoms irrespective of the level of the mediated pathways. More examples specific to policy analysis can be found in Pearl (2009). The concept of a controlled **IE** does not generally exist except in very specific situations (T. J. VanderWeele, 2010) and even under these specific situations, the interpretation of a controlled **IE** has yet to be widely adopted. The focus of this thesis would be on natural **DE** and **IE** effects.

Going back to the definition of **DE**, the words 'pure' and 'total' in the names refer to the predetermined level of r_M in $Y(r_Y, M(r_M))$ when defining the **DE**s. 'Pure' refers to setting R_M

to $r_M = 0$, i.e. the value of M under the control condition $r_M = 0$. 'total' refers to setting the value of R_M to $r_M = 1$, i.e. the value of M under the experimental condition. The same applies to **IE**s where there are *pure natural indirect effect (**PNIE**)* and *total natural indirect effect (**TNIE**)* defined as follows:

$$\textbf{PNIE} \triangleq \mathrm{E}[Y(0, M(1))] - \mathrm{E}[Y(0, M(0))]$$
$$\textbf{TNIE} \triangleq \mathrm{E}[Y(1, M(1))] - \mathrm{E}[Y(1, M(0))]$$

Here, however, the 'pure' and 'total' refer to the level of r_Y in $Y(r, M(M))$ rather than r_M as in the case of the **DE**. In each of the **DE** and **IE** definitions, two **PO**s are used and each **PO** has two values of R. For each definition, one of the R is the same in both **PO** while the other is different. For the **DE**, since the effect of interest does not pass through the mediator, therefore, the mediator in both of the **PO** used in each of the definition is held at the same value of R had M been exposed to the said level. For the **IE**, the effect of interest must pass through the mediator and thus the R value that is held the same is the R under which Y would have been had it been exposed to the said level and the R under M is different for the two **PO**. The 'pure' and 'total' terms then refer to the R that is held at the same value in both **PO**.

Having stated the possible ways to define the **DE** and **IE**, the question remains as to which one to adopt for the current thesis. Various researchers have adopted different approaches over which **DE** and **IE** to use. Pearl (2012) used only **PNDE** and **PNIE** as it was deemed to be easier to interpret the causal effects since the causal effects were assessed under the control condition. Imai et al. (2010) on the other hand defined the **DE** as the mean between the **PNDE** and the **TNDE** and the **IE** as the mean between the **PNIE** and **TNIE**. Under conditions where there are no interactions between the treatment and the mediator, the **PNDE** is the same as **TNDE**. The same is true for the **IE**. However, when there are interactions, the **PNDE** would not equal to the **TNDE** and likewise for the **IE**. This difference motivated the adoption of using the mean between the 'pure' and 'total' effects for both the **DE** and **IE** by Imai et al. (2010).

The next approach, which is also the approach adopted for this thesis was adopted by several researchers including but not limited to Valeri and VanderWeele (2013), Daniel et al. (2015) and Muthén et al. (2016). This approach uses either the **PNDE** with the **TNIE** or the **TNDE** with

the **PNIE** which give rise to the following relationships.

$$\begin{aligned}\textbf{TE} &\triangleq \textbf{PNDE} + \textbf{TNIE} \\ &\triangleq \textbf{TNDE} + \textbf{PNIE}\end{aligned}$$

This approach was adopted because of the intuitive interpretation of the **TE** as the sum of the **DE** and **IE**. Its weakness is that different estimates for the **DE** and **IE** will be obtained if there are any interactions between the treatment and mediator. However, this thesis will only consider scenarios without interactions and the pair of effects that are used are the **PNDE** and **TNIE**. They will henceforth be referred as the **DE** and **IE**, dropping the 'pure' and 'total' labels. As a summary, the definitions of the **DE** and **IE** and their relationship with the **TE** is a follows,

$$\begin{aligned}\textbf{PNDE} &\triangleq \mathrm{E}[Y(1, M(0))] - \mathrm{E}[Y(0, M(0))] \\ \textbf{TNIE} &\triangleq \mathrm{E}[Y(1, M(1))] - \mathrm{E}[Y(1, M(0))] \\ \textbf{TE} &\triangleq \textbf{PNDE} + \textbf{TNIE}\end{aligned}$$

With the **DE** and **IE** defined, we will next consider this issue for binary outcomes where causal **OR**s are used.

Mediated causal OR

We have so far defined mediated causal effects for continuous outcome measures. The same principles apply to effect size measures used for binary outcomes. Using nested **PO**s the **TOR** can be written as

$$\textbf{TOR} \triangleq \frac{\Pr(Y(1, M(1)) = 1)}{\Pr(Y(1, M(1)) = 0)} \Big/ \frac{\Pr(Y(0, M(0)) = 1)}{\Pr(Y(0, M(0)) = 0)}$$

Just like the **TE** where the causal effect is defined as a difference and can be decomposed into a **DE** and an **IE**, the **TOR** also possess a similar decomposition. Following the use of 'pure' effects for **DE** and 'total' effects for **IE** previously, we would use 'pure' and 'total' effects to define the direct odds ratio (**DOR**) and indirect odds ratio (**IOR**) respectively. The **DOR** and **IOR** and their relationship with the **TOR** are defined as follows.

$$\begin{aligned}\textbf{DOR} &\triangleq \frac{\Pr(Y(1, M(0)) = 1)}{\Pr(Y(1, M(0)) = 0)} \Big/ \frac{\Pr(Y(0, M(0)) = 1)}{\Pr(Y(0, M(0)) = 0)} \\ \textbf{IOR} &\triangleq \frac{\Pr(Y(1, M(1)) = 1)}{\Pr(Y(1, M(1)) = 0)} \Big/ \frac{\Pr(Y(1, M(0)) = 1)}{\Pr(Y(1, M(0)) = 0)} \\ \textbf{TOR} &\triangleq \textbf{DOR} \cdot \textbf{IOR}\end{aligned}$$

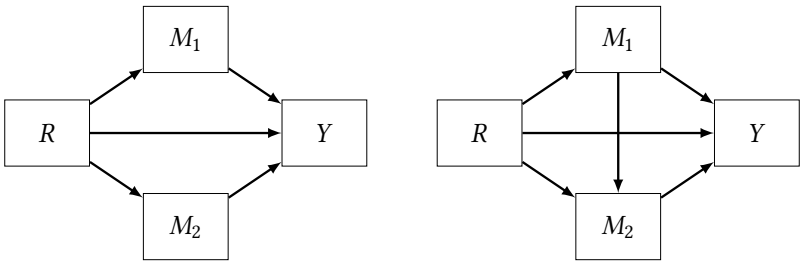
The difference between the relationship of **TE** and its direct and indirect effects with **TOR** and its direct and indirect **OR** is that in the former, the relationship is additive while in the latter, the relationship is multiplicative. Now that the causal effects for continuous (**TE**, **DE**, **IE**) and binary (**TOR**, **DOR**, **IOR**) outcomes for the single mediator case have been defined, I will next move on to the definitions of the same for the two mediators case.

3.4.3 Mediated causal effects for two sequential mediators

In the previous subsection, I defined the different causal effects for a single mediator for both continuous and binary outcomes. In this subsection, I will define the causal effects for two mediators and discuss some of the problems encountered when extending the definitions to accommodate two mediators. The work presented here focused on the two mediator scenario but could be adapted to more generally apply to any number of mediators. It should be noted though that as the number of mediators increase, so does the complexity of the definitions and the estimation of the causal effects which will be discussed in the next section. The increased complexity may make it difficult to interpret what each of the effects precisely mean. A detailed treatment of this and other issues related to multiple mediators can be found in Daniel et al. (2015).

Before going into the definitions, I would like to highlight two different ways of accommodating two mediators, parallel and sequential mediation. Figure 3.5 will be used as a reference to highlight the similarities and differences between the two mediation models. R, as before represents the treatment variable, M_1 represents the first mediator, M_2 represents the second mediator and Y represents the outcome of interest. Arrows within the figure indicate causal relationships. Both models are similar in that R, as the only independent variable in the model, causes changes in M_1, M_2 and Y. The difference between them is the single arrow from M_1 to M_2 and the implication of this is that M_2 is dependent on both R and M_1. This extra dependency of M_2 creates extra paths by which the effect of the treatment R can traverse and also implies that there will be more **PO**s of M_2 than if it was only dependent on R. This complexity is carried forward when assessing the paths through which R can have an effect on Y. This thesis will focus on the definition and estimation of sequential mediation.

Figure 3.5: Mediation diagram with 2 mediators



Parallel mediation Sequential mediation

Exploring this complexity a little further and using fig. 3.5 as a reference, we can enumerate the possible paths from R to Y. These paths are named **DE**, indirect effect through M_1 only ($\mathbf{IE}_1$), indirect effect through M_2 only ($\mathbf{IE}_2$) and indirect effect through M_1 and M_2 ($\mathbf{IE}_3$).

$$\mathbf{DE}: R \rightarrow \quad Y$$

$$\mathbf{IE}_1: R \rightarrow \quad M_1 \rightarrow \quad Y$$

$$\mathbf{IE}_2: R \rightarrow \quad M_2 \rightarrow \quad Y$$

$$\mathbf{IE}_3: R \rightarrow \quad M_1 \rightarrow \quad M_2 \rightarrow \quad Y$$

These four different paths represent the direct and indirect effects of interest. After identifying the effects of interest, they need to be defined using **PO**s. Let us first consider the number of **PO** of M_1, M_2 and then Y. M_1 is dependent on only R, therefore it has only 2 **PO**s.

$$M_1(0), \quad M_1(1)$$

M_2 is dependent on both R and M_1 resulting in a total of 4 **PO**s:

$$M_2(0, M_1(0)), \quad M_2(0, M_1(1)),$$
$$M_2(1, M_1(0)), \quad M_2(1, M_1(1))$$

Lastly, Y depends on R, M_1 and M_2 resulting in a total of 16 **POs**.

$$
\begin{aligned}
&Y(0, M_1(0), M_2(0, M_1(0))), \quad Y(0, M_1(0), M_2(0, M_1(1))),\\
&Y(0, M_1(0), M_2(1, M_1(0))), \quad Y(0, M_1(0), M_2(1, M_1(1))),\\
&Y(0, M_1(1), M_2(0, M_1(0))), \quad Y(0, M_1(1), M_2(0, M_1(1))),\\
&Y(0, M_1(1), M_2(1, M_1(0))), \quad Y(0, M_1(1), M_2(1, M_1(1))),\\
&Y(1, M_1(0), M_2(0, M_1(0))), \quad Y(1, M_1(0), M_2(0, M_1(1))),\\
&Y(1, M_1(0), M_2(1, M_1(0))), \quad Y(1, M_1(0), M_2(1, M_1(1))),\\
&Y(1, M_1(1), M_2(0, M_1(0))), \quad Y(1, M_1(1), M_2(0, M_1(1))),\\
&Y(1, M_1(1), M_2(1, M_1(0))), \quad Y(1, M_1(1), M_2(1, M_1(1)))
\end{aligned}
$$

As in the one mediator scenario, out of the 16 **POs**, only two are observable with the rest being cross-world **POs**. These two observable **POs** can also be alternatively expressed as $Y(0)$ and $Y(1)$ respectively.

$$
\begin{aligned}
Y(0) &= Y(0, M_1(0), M_2(0, M_1(0)))\\
Y(1) &= Y(1, M_1(1), M_2(1, M_1(1)))
\end{aligned}
$$

The causal effects were defined using these 16 Y **POs**. Also similar to the single mediator scenario where there are 'pure' and 'total' effects, in the two mediators scenario, there are multiple ways to define each of the direct and indirect effects. However, unlike the single mediator case, instead of having only two ways to define each effect, there are now eight ways to define each effect. Using **DE** as an example, a list of all the valid definitions can be found in table 3.1. As in the single mediator case, the direct effect is the effect represented by '$R \rightarrow Y$' and therefore the causal effect should reflect the difference between the $r = 1$ and $r = 0$ of the Y **PO** while holding the rest of the r at a natural level. Since there are three other rs within each **PO** and each r can assume either 0 or 1, there are therefore 2^3 or 8 different ways to hold the rest of the **PO** at a natural level. The same applies for each of the indirect effects making it a total of 32 direct and indirect effects. If as before, we were to define the **TE** as a sum of the direct and indirect effects, then out of these 32 effects, there would be 24 combinations of direct and indirect effects that fulfil this condition (Daniel et al., 2015).

Assuming that there were no treatment-mediator and mediator-outcome interaction effects as we had done in the single mediator case, each of the 8 direct effects definitions would produce

Table 3.1: List of valid direct effect definitions

$\mathrm{E}[Y(1, M_1(0), M_2(0, M_1(0)))] - \mathrm{E}[Y(0, M_1(0), M_2(0, M_1(0)))]$
$\mathrm{E}[Y(1, M_1(0), M_2(0, M_1(1)))] - \mathrm{E}[Y(0, M_1(0), M_2(0, M_1(1)))]$
$\mathrm{E}[Y(1, M_1(0), M_2(1, M_1(0)))] - \mathrm{E}[Y(0, M_1(0), M_2(1, M_1(0)))]$
$\mathrm{E}[Y(1, M_1(0), M_2(1, M_1(1)))] - \mathrm{E}[Y(0, M_1(0), M_2(1, M_1(1)))]$
$\mathrm{E}[Y(1, M_1(1), M_2(0, M_1(0)))] - \mathrm{E}[Y(0, M_1(1), M_2(0, M_1(0)))]$
$\mathrm{E}[Y(1, M_1(1), M_2(0, M_1(1)))] - \mathrm{E}[Y(0, M_1(1), M_2(0, M_1(1)))]$
$\mathrm{E}[Y(1, M_1(1), M_2(1, M_1(0)))] - \mathrm{E}[Y(0, M_1(1), M_2(1, M_1(0)))]$
$\mathrm{E}[Y(1, M_1(1), M_2(1, M_1(1)))] - \mathrm{E}[Y(0, M_1(1), M_2(1, M_1(1)))]$

the same estimates. The same holds true for the indirect effects. However, if there are interaction effects, then the different effect definition for each of the direct and indirect effect would produce different results. A possible solution to this is to adopt Imai et al. (2010)'s definition of direct and indirect effects to average across the different possible definitions. This however makes the estimation of the causal effects much more complex so for this thesis, I will focus on the scenario where there are no interaction effects and for that, I have adopted 1 of the 24 combinations of direct and indirect effects shown in eq. (3.4.7). For reference, the combination of causal effects adopted corresponds to decomposition one from Daniel et al. (2015).

$$
\begin{aligned}
\mathbf{DE} &\triangleq \mathrm{E}[Y(1, M_1(0), M_2(0, M_1(0)))] - \mathrm{E}[Y(0, M_1(0), M_2(0, M_1(0)))] \\
\mathbf{IE}_1 &\triangleq \mathrm{E}[Y(1, M_1(1), M_2(0, M_1(0)))] - \mathrm{E}[Y(1, M_1(0), M_2(0, M_1(0)))] \\
\mathbf{IE}_2 &\triangleq \mathrm{E}[Y(1, M_1(1), M_2(1, M_1(0)))] - \mathrm{E}[Y(1, M_1(1), M_2(0, M_1(0)))] \\
\mathbf{IE}_3 &\triangleq \mathrm{E}[Y(1, M_1(1), M_2(1, M_1(1)))] - \mathrm{E}[Y(1, M_1(1), M_2(1, M_1(0)))] \\
\mathbf{TE} &\triangleq \mathbf{DE} + \mathbf{IE}_1 + \mathbf{IE}_2 + \mathbf{IE}_3
\end{aligned}
\tag{3.4.7}
$$

Causal **OR** can also be defined for binary Y in the two mediator case as follows.

$$\mathbf{DOR} \triangleq \frac{\Pr(Y(1,M_1(0),M_2(0,M_1(0)))=1)}{\Pr(Y(1,M_1(0),M_2(0,M_1(0)))=0)} \Big/ \frac{\Pr(Y(0,M_1(0),M_2(0,M_1(0)))=1)}{\Pr(Y(0,M_1(0),M_2(0,M_1(0)))=0)}$$

$$\mathbf{IOR}_1 \triangleq \frac{\Pr(Y(1,M_1(1),M_2(0,M_1(0)))=1)}{\Pr(Y(1,M_1(1),M_2(0,M_1(0)))=0)} \Big/ \frac{\Pr(Y(1,M_1(0),M_2(0,M_1(0)))=1)}{\Pr(Y(1,M_1(0),M_2(0,M_1(0)))=0)}$$

$$\mathbf{IOR}_2 \triangleq \frac{\Pr(Y(1,M_1(1),M_2(1,M_1(0)))=1)}{\Pr(Y(1,M_1(1),M_2(1,M_1(0)))=0)} \Big/ \frac{\Pr(Y(1,M_1(1),M_2(0,M_1(0)))=1)}{\Pr(Y(1,M_1(1),M_2(0,M_1(0)))=0)}$$

$$\mathbf{IOR}_3 \triangleq \frac{\Pr(Y(1,M_1(1),M_2(1,M_1(1)))=1)}{\Pr(Y(1,M_1(1),M_2(1,M_1(1)))=0)} \Big/ \frac{\Pr(Y(1,M_1(1),M_2(1,M_1(0)))=1)}{\Pr(Y(1,M_1(1),M_2(1,M_1(0)))=0)}$$

$$\mathbf{TOR} \triangleq \mathbf{DOR} \cdot \mathbf{IOR}_1 \cdot \mathbf{IOR}_2 \cdot \mathbf{IOR}_3$$

This section ends with the definition of causal effects. In the next section, we turn our focus to estimating these effects.

3.5 Parametric estimation of mediated causal effects

In the previous section, the direct and indirect causal effects for continuous and binary Y have been defined for both the one mediator and two mediators scenario. In this section, we shall discuss how these effects can be estimated using fully parametric models. Broadly, there are two main ways of estimation: parametric and non-parametric estimation. Parametric estimation refers to the use of parametric models where assumptions are made about the distribution of the dependent variable as well as about the functional form of the relationships between variables. Non-parametric estimation does not have such assumptions. Before proceeding, it should be noted that assumptions refer to certain properties that cannot always be verified from empirical data. These assumptions can arise from many sources including the study design used, constraints imposed by the parametric model or associated with a framework such as the causal inference framework adopted in this thesis. Some of these assumptions can be checked using empirical data although in most instances, these checks are only indicative of whether gross violations of the assumptions had taken place. An approach for considering the impact of a violation of some these assumptions will be covered in chapter 6.

We will first turn our attention to non-parametric formulations of estimating mediated causal effects. Although this thesis focuses on parametric estimation of mediated causal effects, under-

standing how non-parametric estimation of mediated causal effects work can provide valuable insight into how the gaps in estimating mediated causal effects can be bridged. Let us first re-examine a mediated causal effect. The definition of **IE** is reproduced here.

$$\mathbf{IE} \triangleq \mathrm{E}[Y(1, M(1))] - \mathrm{E}[Y(1, M(0))]$$

From this definition, we can see that the building blocks for estimating causal effects are the ability to estimate the expectations of nested **PO**s such as $\mathrm{E}[Y(1, M(1))]$.

Using an **RCT** as an example, we first consider how a non-nested **PO** can be estimated. In a typical 2-arm **RCT**, we have an intervention of interest, a control intervention and an outcome of interest. Since there are two interventions, there would also be only two **PO**. The expectation of these two **PO** are $\mathrm{E}[Y(1)]$ and $\mathrm{E}[Y(0)]$, one for each intervention. These two expectations represent the average outcome had everyone been offered the intervention of interest ($R = 1$) or the control intervention ($R = 0$). As discussed earlier, the expectations $\mathrm{E}[Y(1)]$ and $\mathrm{E}[Y(0)]$ can be estimated by taking the mean of all those offered $R = 1$ and $R = 0$ respectively.

Moving on to the estimation of the expectation of a nested **PO** such as $Y(1, M(0))$. The nested **PO** $Y(1, M(0))$ refers to the outcome Y had R been set to the value of 1 and the value of M set to the value of $M(0)$. $M(0)$ refers to the mediator M had R been set to the value of 0. Estimating the expectation such a nested **PO** therefore requires one to first be able to estimate $M(0)$ so that the value of $M(0)$ can be used to set the value of M when estimating $\mathrm{E}[Y(1, M(0))]$. However, unlike the case of a non-nested **PO**, the value of $M(0)$ has a distribution of its own and when nested within the expectation of Y, an estimation of the expectation of Y needs to accommodate this distribution of M.

The solution to this nested expectation derives from a result in statistics known as the *law of iterated expectation* (Billingsley, 1995). An application of the result specifically for the estimation of nested **PO** had been developed and simplified and was named the *mediation formula* (Pearl, 2001, 2012).

Mediation formula

The mediation formula is a series of formulae that uses the *law of iterated expectations* to obtain an expression that aids in the estimation of the expectation of a nested **PO**. The *law of iterated expectations*, also known as the *tower rule* states that for two random variables Y and X, $\mathrm{E}[Y]$,

the marginal expectation of Y can be expressed as follows (Billingsley, 1995):

$$\mathrm{E}[Y] = \mathrm{E}[[\mathrm{E}[Y|X]]$$

The equation is an application of the *law of iterated expectations* and it indicates how the marginal expectation($\mathrm{E}[Y]$) is related to the conditional expectation ($\mathrm{E}[Y|X]$). The right-hand side (**RHS**) of the equation contains a nested expectation and this nesting is a similar scenario as the nesting of a **PO** within another **PO**. What this equation indicates is a way by which we can obtain an estimate for a marginal expectation of a random variable from the its conditional expectation. This in turn suggests that the nested **PO** can be estimated using the same methods as those used to estimate nested expectations using the *law of iterated expectations*. I would use a special case of this law to demonstrate how it can be used. This special case arises when X is finite and the possible outcomes of X are $x_1, x_2, \ldots, x_i$. If we want to estimate the marginal expectation of Y from the conditional expectation of Y on X, we can apply the law and obtain the following expression:

$$\mathrm{E}[Y] = \sum_{i=1}^{x_1,x_2,\ldots,x_i} \mathrm{E}[Y|x_i] \cdot \Pr(x_i)$$

This expression, when evaluated, provides us with an estimate of the marginal expectation of Y. This is a special case because X is discrete and the probability associated with each X can be estimated. These estimated probabilities of the discrete Xs is then used to weight each conditional expectation of Y. These weighted conditional expectations of Y can then be summed to obtain an estimate of the marginal expectation of Y. This ability to estimate the marginal expectation using the conditional expectations together with their corresponding probabilities is what allows us to use the result from the *law of iterated expectations* to estimate the expectations of nested **PO**. Applying this to a nested **PO** with a finite mediator, we will replace the X with M and use the possible values of M in place of possible values of X. This is applicable to mediation but instead of X, we have a random variable M, the mediator, which is conditioned on a fixed value of R. For example, if we want to estimate the nested **PO** $\mathrm{E}[Y(1, M(0))]$, we can apply the result and express the **PO** as follows.

$$\mathrm{E}[Y(1, M(0))] = \sum_{m} \mathrm{E}[Y|R = 1, M = m] \cdot \Pr(M = m|R = 0)$$

For each unique M, we will obtain a value for $\mathrm{E}[Y|R = 1, M = m]$ which is weighted by the probability of $M = m$ conditional on $R = 0$ occurring. The sum of these values will provide an

estimate for the expectation of the nested **PO** $Y(1, M(0))$. In the case of a continuous M, the same idea applies but instead of a summation, it will be an integration term with respect to M and the probability of $M = m$ conditional on $R = 0$ will be replaced by the probability density function (**PDF**) of M.

$$\mathrm{E}[Y(1, M(0))] = \int_m \mathrm{E}[Y|R = 1, M = m] \cdot \Pr(M = m|R = 0)dm$$

Crucially, in both formulations of estimates of the expectation of a nested **PO**, there are no assumed distributions for either Y or M. This non-parametric formulation therefore lends itself for adaptation to a parametric formulation which we will discuss in the next section. The mediation formula essentially applies the law of iterated expectation onto the two **PO**s within each definition of the **DE** and **IE**. Doing so yields the following results,

$$\begin{aligned}
\textbf{DE} &\triangleq \mathrm{E}[Y(1, M(0))] - \mathrm{E}[Y(0, M(0))] \\
&= \sum_m [\mathrm{E}[Y|R = 1, M = m] - \mathrm{E}[Y|R = 0, M = m]] \cdot \Pr(M = m|R = 0)] \qquad (3.5.1) \\
\textbf{IE} &\triangleq \mathrm{E}[Y(1, M(1))] - \mathrm{E}[Y(1, M(0))] \\
&= \sum_m \mathrm{E}[Y|R = 1, M = m] \cdot [\Pr(M = m|R = 1) - \Pr(M = m|R = 0)] \qquad (3.5.2)
\end{aligned}$$

This provides a way forward to estimate each of these effects and from these results, we have identified the key components necessary to estimate each of the effects, i.e. namely the probability of M and the expectation of Y conditional on a set of predetermined parameters. In the next section, I will discuss how these results can be used together in parametric models to estimate each of these causal effects of interest.

3.5.1 Parametric models

Parametric model in the context of this thesis refers to models primarily from the **GLM** family. For continuous and binary dependent variables, the normal **GLM** with identity link and the logistic **GLM** with logit link will be used respectively. In the case of a mediation model, the parametric model is used to model the outcome of interest, Y, and the mediator, M. The use of parametric models mean that each of the components required to estimate the expectation of a nested **PO** can now be easily estimated since there are underlying assumed distributions of components of the Y and M models. Furthermore, parametric models also allow covariates to

be part of the model for Y and M, providing a way by which adjustments to the causal effect could be made if it was deemed necessary.

Normal GLM

The normal **GLM** with identify link for modelling continuous dependent variables is of the form,

$$Y = f(x) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Y represents the outcome of interest and is the dependent variable. X represents the set of explanatory variables and ε represents the error term which is assumed to follow a normal distribution with mean, μ, of 0 and an unknown variance, σ^2. β_0 is also known as the intercept of the model and represents the conditional mean of Y. The unknown βs including β_0 and σ^2 will be estimated using sample data.

Logistic GLM

The logistic **GLM** with logit link for modelling binary dependent variables has a more complex form than its linear counterpart. Before going into how the binary dependent variables are modelled using the logistic regression, it needs to be emphasised that the formulation of the logistic regression used in this thesis is the latent variable formulation. In some representations, the error term of the logistic model, ε, is absent but in the latent variable formulation it is present. The latent variable formulation was chosen for this work because of the explicit representation of ε which enables subsequent development of a sensitivity analysis procedure.

The logistic model is more complex because rather than modelling the binary variable, the probability of each response being a 1 is modelled instead. Since a probability of an event is a bounded number between 0 and 1, it is necessary to transform it to a form that is unbounded to facilitate modelling the probability. The transformation adopted in the logistic **GLM** is the logit transformation. The logit and its inverse, the expit, are given as follows,

$$\text{logit}(x) = \log\left(\frac{x}{1-x}\right)$$
$$\text{expit}(x) = \frac{1}{1+e^{-x}}$$

Using the logit, the logistic **GLM** is of the form,

$$Y = \mathrm{I}((Y^* + \varepsilon) \geqslant 0), \quad \varepsilon \sim \mathcal{L}_{\mathbf{STD}} \tag{3.5.3}$$

$$\begin{aligned} Y^* &= \mathrm{logit}(\Pr(Y = 1)) \\ &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \end{aligned} \tag{3.5.4}$$

The model Y is modelled using a latent variable Y^* which is conceptualised as a logit transformation of the underlying probability of $Y = 1$, $\Pr(Y = 1)$. Y^* is evaluated using an indicator function 'I' where if Y^* is greater than or equal to zero, it evaluates to 1 and to 0 otherwise. X, like before represents the explanatory variables and ε is the error term and follows a standard logistic distribution, $\mathcal{L}(\mu = 0, \sigma^2 = \frac{\pi^2}{3})$. A prediction of Y (otherwise also known as realisation of Y) is done by first making a draw of ε and the passing it through the indicator function. Note that ε does not appear unless it is necessary to make a prediction of Y and this is due primarily because for a binary outcome, we are modelling the underlying distribution of $\Pr(Y = 1)$ and not Y directly. The model allows us to estimate $\Pr(Y = 1)$ but in order to make a prediction of Y which is a binary variable, we have to make a draw of Y using its estimated probability. The process of adding ε to Y^* and assigning one of the two binary values depending on if its above or below zero is the same as making a single draw of 0, 1 with the probabilities $\{1 - \mathrm{expit}(Y^*), \mathrm{expit}(Y^*)\}$. There is however one advantage however of using the method of adding the ε and passing it through the indicator function. This advantage will be elaborated in the next chapter in the discussion of using simulations to estimate mediated causal effects.

Applying the GLMs to the mediation models

For a single mediator continuous Y and M would be modelled using the following models.

$$Y = \gamma_0 + \gamma_1 R + \gamma_2 M + \gamma_3 C + \varepsilon_Y, \quad \varepsilon_Y \sim \mathcal{N}(0, \sigma^2) \tag{3.5.5}$$

$$M = \beta_0 + \beta_1 R + \beta_2 C + \varepsilon_M, \quad \varepsilon_M \sim \mathcal{N}(0, \sigma^2) \tag{3.5.6}$$

where R = intervention offered

Y = outcome

M = mediator

C = covariate

$\varepsilon_Y, \varepsilon_M$ = error terms

For simplicity, C is presumed to be a single covariate and is the same for both the mediator and outcome models but it could be multiple covariates and be different for the mediator and outcome models. Binary Y and M would be modelled using logistic regression as follows.

$$\begin{aligned} Y &= \mathrm{I}(\mathrm{logit}(\Pr(Y=1))) \\ &= \mathrm{I}(\gamma_0 + \gamma_1 R + \gamma_2 M + \gamma_3 C + \varepsilon_Y) \qquad (3.5.7) \\ M &= \mathrm{I}(\mathrm{logit}(\Pr(M=1))) \\ &= \mathrm{I}(\beta_0 + \beta_1 R + \beta_2 C + \varepsilon_M) \qquad (3.5.8) \end{aligned}$$

where R = intervention offered

Y = outcome

M = mediator

C = covariate

$\varepsilon_Y \sim \mathcal{L}_{\mathrm{STD}}$

For two sequential mediators, it would simply be an extension of what was done for the single mediator. For continuous Y, M_1 and M_2, they would be modelled using the following models.

$$Y = \gamma_0 + \gamma_1 R + \gamma_2 C + \gamma_3 M_1 + \gamma_4 M_2 + \varepsilon_Y \qquad (3.5.9)$$

$$M_1 = \beta_0 + \beta_1 R + \beta_2 C + \varepsilon_{M_1} \qquad (3.5.10)$$

$$M_2 = \alpha_0 + \alpha_1 R + \alpha_2 C + \alpha_3 M_1 + \varepsilon_{M_2} \qquad (3.5.11)$$

where R = intervention offered

Y = outcome

M = mediator

C = covariate

$\varepsilon_Y \sim \mathcal{N}(0, \sigma^2)$

For binary Y, M_1 and M_2, they would be modelled using the following models.

$$\begin{aligned} Y &= \mathrm{I}(\mathrm{logit}(\Pr(Y=1))) \\ &= \mathrm{I}(\gamma_0 + \gamma_1 R + \gamma_2 C + \gamma_3 M_1 + \gamma_4 M_2 \varepsilon_Y) \qquad (3.5.12) \\ M_1 &= \mathrm{I}(\mathrm{logit}(\Pr(M_1=1))) \\ &= \mathrm{I}(\beta_0 + \beta_1 R + \beta_2 C + \varepsilon_{M_1}) \qquad (3.5.13) \\ M_2 &= \mathrm{I}(\mathrm{logit}(\Pr(M_2=1))) \\ &= \mathrm{I}(\alpha_0 + \alpha_1 R + \alpha_2 C + \alpha_3 M_1 + \varepsilon_{M_2}) \qquad (3.5.14) \end{aligned}$$

where R = intervention offered

Y = outcome

M = mediator

C = covariate

$\varepsilon_Y \sim \mathcal{L}_{\mathbf{STD}}$

Relationship between parametric models and causal effects.

After fitting an appropriate parametric model for Y, M_2 and M_1, we can then proceed with the estimation of the causal effects. maximum likelihood estimation (**MLE**) is used to fit the models from which we then extract the maximum likelihood estimators for each of the unknown model parameters. The next section will discuss the various forms of estimation methods for the causal estimand using these extracted maximum likelihood estimators.

With each of the possible dependent variables stated, we now move to how a causal effect can be estimated. Recall that to estimate any of the causal effect, we need to be able to estimate two nested **PO** and then use the **PO** to compute the causal effect either by taking a difference in the case of continuous outcomes or taking the **OR** in the case of binary outcomes.

Sticking to the case of the continuous outcomes first, recall that we can estimate the **DE** and **IE**

by using the mediation formula as follows,

$$\begin{aligned}\mathbf{DE} &\triangleq \mathrm{E}[Y(1, M(0))] - \mathrm{E}[Y(0, M(0))] \\ &= \sum_m [\mathrm{E}[Y|R=1, M=m] - \mathrm{E}[Y|R=0, M=m]] \cdot \Pr(M=m|R=0)] \\ \mathbf{IE} &\triangleq \mathrm{E}[Y(1, M(1))] - \mathrm{E}[Y(1, M(0))] \\ &= \sum_m \mathrm{E}[Y|R=1, M=m] \cdot [\Pr(M=m|R=1) - \Pr(M=m|R=0)]\end{aligned}$$

Since we now have a model for each of Y and M, the general approach to estimating the causal effect is to use the models to provide estimates for the cross-world **PO**. This is vital since these cross-world **PO** can not be observed in reality. These parametric models are a set of equations that summarises what reality is from sample data. This summary is a best guess of the relationships between the explanatory variables and the dependent variables under certain assumptions. Ideally, what we want to use the model for is to be able to ask it questions such as 'if the value of R was set to 1, what will the value of M be?'. Now that we are able to 'ask' the set of models questions for 'what-if' scenarios and obtain estimates that address the question, the last task that needs to be done is how to put together these answers and obtain an estimate of the causal effect of interest and the uncertainty associated with this estimate.

The methods discussed mainly differ by:

1. How to 'ask' the model questions?
2. How do I combine the responses?
3. How do I obtain uncertainty bounds on the estimate?

Starting with the first, I will discuss the problems associated with each.

Obtaining estimates for a nested PO

Given that the parametric models explicitly model the joint conditional distribution of Y on the independent variables, it appears straightforward to obtain estimates for a given set of independent variables as defined in a nested **PO**. Intuitively, it appears to be as simple as substituting the required Xs into the model and obtaining estimates for Y. However, because of the nested nature of the **PO**, doing so will only yield the correct estimate for the case where both the Y and M are continuous. In order to illustrate this, we first consider a continuous Y

and M. Applying the mediation formula to the **PO** $Y(r_Y, M(r_M))$, we would have the following:

$$\begin{aligned}
\mathrm{E}[Y(r_Y, M(r_M))] &= \int_m \mathrm{E}[Y|R = r_Y, M = m] \cdot \Pr(M|R = r_M)dm \\
&= \int_m (\gamma_0 + \gamma_1 r_Y + \gamma_2 M + \gamma_3 C) \cdot \Pr(M|R = r_M)dm \\
&= \gamma_0 + \gamma_1 r_Y + \gamma_3 C + \gamma_2 \int_m M \cdot \Pr(M|R = r_M)dm \\
&= \gamma_0 + \gamma_1 r_Y + \gamma_3 C + \gamma_2 \,\mathrm{E}[M|R = r_M] \\
&= \gamma_0 + \gamma_1 r_Y + \gamma_3 C + \gamma_2(\beta_0 + \beta_1 r_M + \beta_2 C) \qquad (3.5.15)
\end{aligned}$$

The subscript of r differentiates the value r for the outcomes (Y) and mediators (M) under different **PO**. Next, consider a binary Y and a continuous M, we have the following.

$$\begin{aligned}
\mathrm{E}[Y((r_Y, M(r_M))] &= \int_m \mathrm{E}[Y|R = r_Y, M = m] \cdot \Pr(M|R = r_M)dm \\
&= \int_m \Pr(Y = 1|R = r_Y, M) \cdot \Pr(M|R = r_M)dm \\
&= \int_m \Pr(\gamma_0 + \gamma_1 r_Y + \gamma_2 M + \gamma_3 C) \cdot \Pr(M|R = r_M)dm \\
&= \int_m \frac{1}{1 + e^{-[\gamma_0 + \gamma_1 r_Y + \gamma_2 M + \gamma_3 C]}} \cdot \Pr(M|R = r_M)dm \qquad (3.5.16)
\end{aligned}$$

The integral does not yet have a closed form solution and this means that we cannot simply use a formula to obtain an estimate of the expectation of the nested **PO**. It is for this reason that some researchers (e.g. Muthén et al. (2016)) have advocated the use of probit rather than logit as the link functions in the **GLM**. With a probit link function, the error term of the regression is fixed and assumed to follow a normal distribution. Since the outcome and the mediator now share the same error distribution, there is a closed form solution to this integral and hence a formula can be derived to estimating each **PO**. The situation highlighted occurs when a binary dependent variable has a continuous dependent variable as one of its predictor. Specifically, this happens for the following combinations of outcomes and mediators.

1. 1 mediator
    - binary Y, continuous M
2. 2 mediators
    - binary Y, continuous M_1, continuous M_2
    - binary Y, continuous M_1, binary M_2
    - binary Y, binary M_1, continuous M_2
    - binary Y, binary M_1, binary M_2

- continuous Y, continuous M_1, binary M_2

Since the **OR** estimated from a logistic regression is still in widespread use particularly by epidemiologists, researchers have sought different ways to evaluate the integral in eq. (3.5.16). There are two main ways in which the integral can be evaluated. The first is to apply modern computer assisted integration techniques to evaluate the integral. This requires both a high degree of familiarity of the characteristics of the functions to be integrated and the different forms of computer assisted integration to identify suitable techniques to evaluate the integral.

The second method is be able to make draws of the outcome under different **PO** from its underlying distribution. These draws can then be used to simulate the **PO** and this simulated **PO** can then be used to conduct the estimation of the causal effects. This second approach is what I have adopted and it follows closely with the methods adopted by Imai et al. (2010) with some important modifications to allow for the estimation of causal **OR** and sequentially mediated causal effects. Both the computer-assisted integration and the simulation methods are evaluating the same quantity but the simulation method is more adaptable because it can be easily applied to different parametric models which we will look into in greater details in section 3.5.1.

The second problem relates to causal effects for linear and non-linear models. In the case of linear models, the causal effect defined using differences between **PO** does not depend on the covariates. This is not true of non-linear models. This can be illustrated using eq. (3.5.15). Consider the **DE** for continuous Y and M modelled using normal **GLM** as presented in the previous sub-section.

$$\begin{aligned}
\textbf{DE} &\triangleq \mathrm{E}[Y(1, M(0))] - \mathrm{E}[Y(0, M(0))] \\
&= (\gamma_0 + \gamma_1 + \gamma_3 C + \gamma_2(\beta_0 + \beta_2 C)) - (\gamma_0 + \gamma_3 C + \gamma_2(\beta_0 + \beta_2 C)) \\
&= \gamma_0 + \gamma_1 + \gamma_3 C + \gamma_2(\beta_0 + \beta_2 C) - \gamma_0 - \gamma_3 C - \gamma_2(\beta_0 + \beta_2 C)) \\
&= \gamma_1
\end{aligned}$$

In the normal **GLM** case, it can be seen that the because the link function does not do any non-linear transformations, terms that are the same in both **PO** within a causal effect definition get cancelled out. This applies mainly to any covariates present. What this indicates is that the causal effects depend solely on changes in the value of R both when it is in the outcome

model as well as the mediator model. In the case of non-linear regressions such as the logistic regression, the link function links the outcome with the independent variables in a non-linear manner. There is no cancellation of the covariate terms in this case and thus the causal effect estimates depend on both the level of R and the covariates.

In the next section, I will briefly review existing methods for estimating mediated causal effects, focusing on the motivations behind each new method and the general approach taken.

3.5.2 Parametric estimation of mediated causal effects

Mediated causal effects can be estimated in a few ways. The focus in this thesis is on the use of parametric models from the **GLM** family to conduct the estimation. The adoption of **GLM** for the estimation also implies that assumptions associated with the use of parametric **GLM** also applies to the estimation of causal effects. Namely, the independent variables are uncorrelated, each data point is independent and the assumed residuals distribution is correctly specified Additionally, since there are multiple models in use for any given mediation problem, it is further assumed that the error terms of each model are uncorrelated. This additional assumption is related to the idea of a hidden confounder which will be discussed at greater length in chapter 6.

On the methods used to estimate mediated causal effects, two such methods, the *product of coefficients* and *difference in coefficients* methods were used by Baron and Kenny (1986) to estimate statistical mediation. Using the case of a single mediator as an example and assuming that the outcome of interest, Y, and mediator, M, have both been modelled using a normal **GLM**, the *product of coefficients* method can be used to estimate the **IE**. The **IE** in this case is a product of the regression coefficient of R in the M model with the regression coefficient of M in the Y model. The **DE** is the regression coefficient of R in the Y model. The **TE** is the sum of the **DE** and **IE**, consistent with how **TE** had been defined earlier. The **TE** can alternatively be obtained by re-fitting a Y model without the M. Since the **TE** and **DE** can both be obtained directly by fitting different models, and the sum of the **DE** and **IE** is the **TE**, we can obtain the **IE** by subtracting the **DE** from the **TE**. This is known as the *difference in coefficients* method.

We have so far discussed the *product of coefficients* and difference in coefficients methods in the context of a continuous Y and M. The two methods, while applicable to continuous Y

and M, are not applicable when either or both are binary. Extensions to the two methods have been developed to allow Y to be binary (and modelled using logistic **GLM**) and M to be continuous (MacKinnon, 2008). These extensions however require additional assumptions such as uncorrelated error terms for all models and no interactions between all variables. More importantly however is that the estimates arising from these extensions are not consistent between the two methods i.e. the **IE** estimate using the *product of coefficients* method can be different from that estimated using the *difference in coefficients* method (Pearl, 2012). Additional qualifications to the use of each method makes it difficult to generalise the methods for a wide range of scenarios.

A different approach at obtaining causal effect estimates from a binary Y and a continuous M was proposed by Valeri and VanderWeele (2013). This approach, framed within the **PO** framework and based upon the mediation formula sets out to provide a general method to estimate mediated causal effects for a binary Y and continuous M. As discussed previously, if we apply the mediation formula for a binary Y and a continuous M, we will end up with a formula for an estimate of the causal effect that includes an integral term. There is currently no known way to solve for this integral in an exact manner and therefore, no causal effect estimate can be made without some way to solve for this integral. An approximation developed by Valeri and VanderWeele (2013) overcomes the limitations posed by this integral but the proposed approximation requires that the event rate for Y to be low, generally taken to mean below 5%. Gaps therefore still exists for scenarios for higher event rates for instances where a continuous mediator is nested within the outcome or another mediator.

Towards a general method of estimation

Since one of the main challenges of estimation of mediated causal effects was the evaluation of an integral, this challenge became more tractable with improvements in computing power of personal computers as well as advances in the use of software aided integration. There were two main ways in which this challenge could be overcome. The first was the use of numerical integration or otherwise also known as numerical quadrature (Monahan, 2011). These methods were able to approximate the integrals to a high level of precision but they require knowledge of the functional form of the integral. Different functional forms can have different optimum strategies in evaluating the integral and just as different strategies are used to solve integrals, the same applies to computer-aided integration with each method having a trade-off between

precision and computational resources and time required. The complexity introduced by this method of integration namely the requirement of the knowledge of the functional form of the integral as well as the optimum integration methods to apply makes this method of integration difficult to generalise to a wide class of problems.

The second method of estimating an integral is through the use of simulations. This method evaluates the integral by sampling from the underlying distribution of the integral. This additionally allows estimations of integrals with nested distributions as encountered in nested **PO**. Several variations of this simulation method exists and the method adopted by this thesis is a quasi-Bayesian simulation developed by King et al. (2000). This simulation approach uniquely addresses two different forms of uncertainty inherent in a **GLM** which King et al. (2000) termed *estimation* and *fundamental uncertainty*. *Estimation* as its name suggests is associated with the uncertainty associated with the estimation of the model. This form of uncertainty is often represented as confidence interval bounds on estimates of parameters in a model. Fundamental uncertainty is the uncertainty due to chance or unknown events and is often represented as the unaccounted variance of the dependent variable in a model. This simulation method makes draws from an assumed distribution of the regression coefficient as a means of simulating the estimation uncertainty. The sum-product of a single draw with the dataset represents a single draw from the *prior* distribution. Using the case of a normal **GLM** as an example, the fundamental uncertainty can be simulated by generating a vector of normally distributed random numbers with a mean of zero and an estimated variance corresponding to the residual variance of the **GLM**. This vector is added to the sum-product to simulate the fundamental uncertainty. This simulation method is a quasi-Bayesian method because the prior distribution is derived from the initial estimation of the **GLM** rather than obtained using previously known distributional properties of the dependent variable. This simulation method allows one to build up a distribution of Y which takes into account the different forms of uncertainty associated with its estimation. Furthermore, by setting values of the independent variables, we can obtain distributions of Y under different **PO** and it is this unique capability that makes this simulation method well-suited for simulating **PO**s that we can use to compute the causal effect estimates. This thus indicates that this simulation method can be used to develop a general approach towards estimation of mediated causal effects.

Before discussing how this simulation can be adapted to be a general estimation method, the

following section discusses some of the currently available methods and implementations to estimate mediated causal effects.

3.6 Estimation implementations

In the previous section, various existing estimating methods of mediated causal effects were discussed. This section discusses some of the notable implementations of the estimation methods used. It should be noted that the assumptions of causal inference apply no matter which estimation method is adopted. Namely, they are the *exchangeability*, *positivity* and *consistency* assumptions. Crucially, in each of these implementations, there are no straightforward ways to conduct estimation for sequentially mediated causal effects in the presence of more than one mediator.

Structural equation models

Structural equation model are fully parametric models for evaluating causal relations between variables. Structural equation model shares many similarities with generalised linear model and some important differences. The structural equation model uses **GLM** to conduct estimation of the regression models specified in the model but while a single **GLM** is interested in the relationship between the dependent variable with a set of independent variables within a rigid set of assumptions. One of the important differences between the **GLM** model and an **SEM** model is that while the **SEM** uses **GLM** to conduct much of its modelling, the **SEM** additionally allows one to specify relationships between dependent variables in multiple generalised linear model.

An **SEM** is typically theory driven and often used in the context of a comparison between a highly specified model with one that has less specifications. These comparisons provide insight into whether or not the additional specifications are supported by the data. **SEM** is also commonly used to address questions of causal inference and this is not surprising given its beginnings as a method used by Wright (1934) (Pearl & Mackenzie, 2018) to graphically represent causal relationships. Given its flexibility in specifying models, it is a natural fit for translating beliefs about the causal nature of variables into a form that can be tested statistically. The estimates of the pathways in an **SEM** needs further computation in order to obtain the

causal estimands. Some implementations of **SEM** do it automatically when requested such as *Mplus*, a popular software for fitting **SEM** (Muthén et al., 2019).

In terms of estimation of causal estimands, the **SEM** faces similar problems as highlighted previously, namely, the inability to evaluate the integral encountered when applying the mediation formula. Notably implementations of estimation using **SEM** include *Mplus* (Muthén et al., 2019) and *lavaan* (Rosseel, 2012) with the latter being an *R* package (R Core Team, 2018) and is open-sourced. *Mplus* resolves the issue of evaluating the integral by conducting numerical integration using 104 points across a fixed limit of −5 to +5 (Muthén et al., 2016). Although Mplus appears to be adequate for the purposes of estimating the causal effects of interests, the approach of fixing the integration parameters with no way to change them introduces the uncertainty of whether a given set of limits or the number of points used for estimating the integral are adequate.

Furthermore, while *Mplus* is a piece of highly competent software package for a wide variety of uses, it is not open source. This makes it difficult to ascertain the precise manner in which the parameters were estimated. The authors however attempt to plug this gap by providing a highly responsive forum and providing technical documentations on the implementations of the various estimation methods implemented within. An independent attempt to determine the precise manner in which the parameters are estimated had been made by the author of the lavaan where he implements similar functions within lavaan. The source code of lavaan also extensively documents how *Mplus* conducts the estimation of parameters and explains the rationale when the approach of lavaan deviates from *Mplus* and from the documentation made available by the author of lavaan, it can be seen that some default parameters change over time and unless there is a consistent, concerted effort to document them, each new version of the software does require a round of vetting of the software to ensure that it conducts the estimation according to what was previously specified.

This thesis makes used of *Mplus* in the sensitivity analysis of mediated causal estimands because *Mplus* was the only software package that has the capability to conduct logistic **GLM** using **SEM** and setting up constraints to the parameter estimates.

The same feature within lavaan was not available and it is hoped that as lavaan improves in capability, a next iteration of the methods developed within this thesis can use this to conduct

the estimation.

Paramed

paramed is an implementation of the approach developed by Valeri and VanderWeele (2013) where an approximation was used when Y is binary and M is continuous. In all other combinations of outcomes, i.e. when both Y and M are continuous or binary, and when Y is continuous and M is binary, *paramed* uses the application of the mediation formula outlined by Valeri and VanderWeele (2013). *paramed* is currently available as a *Stata* package (Emsley & Liu, 2013), a *SAS* macro and an *R* package.

gformula

gformula is a fully parametric method developed by Daniel et al. (2011) with the intended application in observational studies with time varying exposures or treatments and time-dependent confounders. At its core are two main components: the *g-computation* (Robins, 1986) procedure and *marginal structural models*. The *g-computation* procedure is a procedure much like quasi-Bayesian approach discussed earlier with one key difference: it does not consider the uncertainties associated with the estimated parameters of the models. Like the quasi-Bayesian approach, the *g-computation* starts with fitted models for each of the dependent variables. It then proceeds by using predictions of these models, substituting the exposure or treatment level to the one corresponding to the **PO**s simulated. These predictions are then used in place of the unobserved or unobservable **PO**s. In other words, the fitted models are used as data generating models for the missing **PO**s.

The **PO**s are then placed within a *marginal structural model* for the estimation of the causal effects. *gformula* is primarily concerned with the estimation of causal effects within longitudinal models. Similar to other estimation methods, *gformula* needs a way to evaluate the aforementioned integral. It does so by using assumed distributions between multiple, repeated observations of the same subject to conduct simulations to evaluate the integral. There are implementations of *gformula* in *R* and *Stata* in the form of user contributed packages.

Medflex

medflex is an *R* package developed by Steen et al. (2017) to address the problem of using non-linear models for mediation. The issue at hand as stated by Steen et al. (2017) is that the causal

effects for non-linear models would give rise to causal effects that are dependent on not just the treatment offered but the covariates, as was discussed previously. This dependency also means that the marginal causal effect estimated from such models depend heavily on the make up of the sample. These causal effect estimates would then only be valid for a population that closely resembles the make up of the sample. Once there are deviations to the make up of the population, it is unclear how the deviations would affect the causal estimates due to the non-linear manner by which the outcome is related to the predictor. One of the goals of *medflex* is therefore to estimate the causal effects with an estimator that has a straightforward relationship between covariates and causal effects to make it easier to understand how the causal effect would change for some change in the covariates. In order to achieve this, the authors developed a class of models that they term *natural effects models*. Natural effect models are a class of *structural mean models* that conditions the mean on the covariates and the causal effects estimated from these models are still stratum specific (Dunn et al., 2015).

The key difference in estimation strategy between *medflex* and other packages is that it has two ways by which to simulate the **PO** (weighting vs. imputation based). Each of which have a different set of assumptions and capability. The last thing of note is that *medflex* approached sequential mediators by modelling them jointly. This makes it difficult to tease out mediator specific contributions to the mediated pathways which is often the goal of **PE**.

mediation

mediation is an *R* package developed by Tingley et al. (2014) based upon the methods outlined in Imai et al. (2010). Mediation allows the estimation of mediated causal effects of parametric models and allows a broad range of models to be used for both the outcome and the mediator. It estimates mediated causal effects using a quasi-Bayesian simulation method which also uses the fitted models as data generating models.

The *mediation* package introduced several novel methods in the estimation of mediated causal effects and I will point out four notable ones, namely, defining causal effects using the average causal direct effects (**ACDE**) and average causal mediated effect (**ACME**), the use of a quasi-Bayesian simulation method which uses fitted parametric models as data-generating constructs, allowing the estimation of mediated effects in multiple mediators and the introduction of a robust framework for sensitivity analysis of mediated causal effects to determine the extent in

which unmeasured confounding exerts its effects on the estimated causal effect.

The first novel aspect of *mediation* lies in its use of *average* direct and indirect effects. Both of the direct and indirect effects are defined as the mean between the *total* and *pure* versions of these effects. The motivation of such a definition is to overcome the necessity to choose between using the *pure* or *total* effects in the scenario where there are interactions.

The second novel aspect relates to the use of a quasi-Bayesian simulation method developed by King et al. (2000). This was discussed previously and will be discussed in greater detail in chapter 4.

The third novel aspect of *mediation* is that it allows one to estimate causal effects for more than a single mediator. However, this analysis is restricted to consideration of the mediators jointly and limits one's ability to test path-specific hypothesis using the package. In the context of **PE**, the ability to test path-specific hypotheses is an important tool in confirming theory. This forms one of the main areas of development in this thesis, to allow the testing of path-specific hypotheses.

The fourth and final novel aspect of *mediation* which is being discussed is the implementation of a robust framework to conduct sensitivity analysis. This includes defining the aims of the sensitivity analysis, what the analysis shows, and laying down the conceptual and methodological methods to conduct sensitivity analysis. This provides the tools for one to consider the implication of unmeasured confounding on the causal estimates.

Lastly, while *mediation* provides for many use cases, there are two important things that the *mediation* does not cater for. Firstly, it does not allow the testing of specific hypothesis in sequential mediation for multiple mediators and secondly, it does not have facilities to estimate the causal **OR**. The lack of an estimator for the causal **OR** also implies that there are no facilities to conduct sensitivity analysis when the **OR** is used as the causal effect.

Confidence intervals

After discussing various methods for estimating mediated causal effects, I will next address how the confidence intervals are to be derived. This discussion on confidence intervals is separate from the discussion on the estimation of mediated causal effects because various methods for deriving confidence intervals can generally be adapted for use with each of the methods for estimating mediated causal effects.

Three different forms of confidence intervals would be discussed. The first is the *delta* method. The *delta* method is the method most commonly used for deriving the confidence intervals and it does so by using the standard errors. One of its most prominent uses is in the derivation of confidence intervals in linear regressions. An important characteristic of confidence intervals derived using the *delta* method is that it is symmetrical. The symmetry of confidence intervals derived using the *delta* method arises from the assumption that the estimated quantity, in our case it is the estimated causal effect, has a normal distribution (MacKinnon, 2008). This assumption is valid for direct effects but when applied to indirect effects as proposed by Sobel (1982) causes an inflation of type II errors because indirect causal effects were found not to be normally distributed (MacKinnon, 2008).

The second confidence interval of interest is the bootstrap (Efron, 1987). Notably, the use of non-parametric bootstrap to generate confidence intervals does not make any assumptions of normality and for confidence intervals for indirect causal effects, bootstrap can be used to generate asymmetrical confidence intervals (Efron & Tibshirani, 1994). Very broadly speaking, one would obtain the confidence intervals using non-parametric bootstrap by estimating the causal effect of interest multiple times, each time with a new sample. Each new sample is obtained by resampling the original sample with replacement. Over many runs, we would form a distribution of the causal effect estimates and the causal estimates can then ranked. If we are interested in the 95% confidence interval, we would simply use the $97.5th$ percentile value of the ranked estimates as the upper bound and the $2.5th$ percentile value as the lower bound for the confidence intervals.

The third confidence of interest is similar to the idea of the bootstrap but instead of conducting resampling, we can use simulates of the causal effects. This is the approach adopted by Imai et al.

(2010) in deriving the confidence intervals for the causal estimates. The simulated causal effects similarly form a distribution of the causal effect of interest and we can obtain the confidence intervals by ranking the causal effects and obtaining the required percentile values from the ranked causal effects. This third method is an adaptation of the bootstrap confidence intervals and has the advantage of not requiring to conduct a bootstrap solely for the purpose of obtaining confidence intervals. This thesis follows the implementation of confidence intervals using simulated outcomes.

3.6.1 Gap in methods for the estimation of OR

This section ends off with highlighting gaps in existing methods. Firstly, there is widespread adoption of the **OR** as a causal estimand even though there are several scenarios that currently do not have existing methods to conduct the causal **OR**. Secondly, there appears to be a lack of methods for the analysis of sequential mediation particularly when the question of interest is related to the specific effect of each mediator in the causal chain. Thirdly, sensitivity analysis for binary outcome models is also lacking. The next chapter focuses on addressing these gaps.

3.7 Summary

This chapter started with a discussion of the **PO** framework and how it is used to define causal effects. The definitions of causal effects were then formally stated for differences and **OR** as well as for the single mediator and two mediator case. Examples from the **CASIS** trial were used to motivate the definitions. The chapter ended with a discussion on how parametric methods are used to estimate mediated causal effects, highlighting the main challenges that must be overcome, the existing solutions to these challenges and finally gaps in methods that would be useful for the analysis of **PE**. The next chapter would detail the proposed method for estimation of mediated causal effects.

From the discussion of the various methods of estimation, there are two main methodological gaps that are not fulfilled by existing methods. The first is the use of the causal **OR** as a causal estimand. The reason for not being more widely used as a causal estimand can be traced to

two reasons. Namely, the difficulty in estimation of causal **OR** and the difficulty in interpreting it. The difficulty in estimation as discussed can be overcome with simulation methods. The problems with interpreting **OR** have been well documented and they are not restricted when conducting mediation. In this sense, if the causal **OR** is deemed to be the most suitable causal estimand either by choice or by convention, then researcher will have to consider the issues of interpreting the causal **OR** carefully when using it. This can only take place only when the causal **OR** can be readily estimated and this is what this thesis sets out to do.

The other gap in methods lies in the estimation of sequential mediation when there is more than one mediator. This can be done with an expansion of the mediation formula to consider an additional mediator and this thesis will set out exactly how this can be done.

Chapter 4

Development of a novel estimator for causal odds ratio and sequentially mediated causal effects

4.1 Introduction

The last chapter ended with a discussion of the existing methods for the estimation of mediated causal effects, and highlighted the gaps present in the current methods of estimations. There are two gaps present in existing literature which I aim to bridge in this thesis. The first gap concerns methods for estimating causal estimands in sequential mediation where there are two or more mediators. While methods currently exist to estimate up to two mediators, for example in the *mediation* package by Imai et al. (2010), the analysis is restricted to a conceptual framework where a single mediator of interest is chosen. All other mediators are then considered 'nuisance' mediators and their impact on the causal effect is assessed collectively. This restricts one's ability to test hypothesis relating to the pathways specific to each of the mediator. There is thus a gap for methods for the estimation of mediated causal effects where the effects due to each mediator can be assessed independently. A novel extension of the mediation formula was developed and used to bridge this gap. The work relating to this extension is discussed in the following section.

After extending the mediation formula to accommodate more than a single sequential mediator,

a second gap still exists. This concerns the lack of an estimator for causal odds ratio (**OR**) that does not rely on restrictive assumptions such as that developed by Valeri and VanderWeele (2013). The focus on causal **OR** is due primarily to its widespread use in clinical and epidemiological research. For this reason, bridging this gap will open up its use in these fields and is one of the aims of this thesis. Achieving this aim again starts with the mediation formula. The mediation formula allowed for a general way to define and express mediated causal effects using the potential outcome (**PO**) framework by first expressing the causal effect of interest using **PO**s (Pearl, 2001). In the context of this thesis, this expression is a difference between **PO**s for continuous outcomes, and odds ratio of **PO**s for binary outcomes. Each of the **PO**s in this expression is then expanded using known properties of nested expectations by applying the law of iterated expectations (Billingsley, 1995). After expansion, the expression is then simplified and for many of these expressions, the simplified form can be used like a formula where you can substitute estimated model parameters of the mediator(s) and outcomes to obtain an estimate of the mediated causal effects. This formula simplified from the mediation formula gave identical estimates as the *product of coefficients* method in instances where the *product of coefficients* can be used, suggesting that they are using similar ways to conduct the estimation.

The simplification of the mediation formula and the similarity of estimates between using the simplified mediation formula and the *product of coefficients* method appear to suggest that a formula can be derived for the causal **OR** using the same approach. This is not the case and the problem lies with the lack of a closed form solution to an integral term from expanding the **PO**s. Since there are no closed form solutions, the integral remains in the final expression of the mediated causal effect as discussed in eq. (3.5.16). Given that the expression for estimating the causal effect contains an integral term, this integral term have to be evaluated in order to obtain an estimate. Notably, Valeri and VanderWeele (2013) developed a formula which approximated this integral and hence allowing a way for the **PO**s to be evaluated, thus providing a way to estimate the mediated causal effect. Due to the way this approximation formula was constructed, it produced valid estimates only when the outcome was rare. This limitation left a gap remaining in the estimation of causal **OR** where the outcome is not rare. An alternative to using this approximation formula was to turn towards computer assisted integration methods. One of these methods, which can be used generally to evaluate integrals, is through simulation. Integration through simulation had been used successfully in methods of Bayesian statistics to 'mix' distributions and the same method can and had been applied to evaluate the integral

discussed previously. This was first used by Imai et al. (2010) in the *mediation R* package to estimate mediated causal effects. The simulation method used in *mediation* was based upon prior work by King et al. (2000) who developed the simulation methods as a means to conduct policy analysis. The problem which King et al. (2000) set out to solve was that after a model had been developed using observed data, the researcher might be interested in how the outcome would change if the independent variables assumed some other values. If one were to simply use the values of the independent variable of interest to predict the outcome, we would obtain an estimate of the outcome. This however, ignores the various kinds of variability represented in the model. The method developed by King et al. (2000) provided a quasi-Bayesian way by which these variability could be accounted for using simulations. Since the estimation of mediated causal effects using **PO**s also involves setting the independent variable to a certain value of interest, Imai et al. (2010) was able to adapt the simulation technique to estimate mediated causal effects and the result of these efforts were implemented in *mediation*.

As discussed in the previous chapter, while *mediation* broke grounds on several fronts, there remain gaps relating to the estimation of mediated causal effects. This thesis thus aims to extend the work started in the *mediation* to bridge the gaps identified. The extension proceeded in two steps. The first was to adapt *mediation* such that it could estimate mediated causal effects for more than a single mediator and the second step was to adapt to allow for the estimation of causal **OR**. This extension faced several challenges primarily due to the non-collapsible nature of the **OR** and this necessitates careful consideration of how causal estimates were to be combined within and between simulations in order to obtain a valid causal **OR** estimator. These considerations also tie in strongly with the need to distinguish between marginal and conditional causal **OR** where the marginal causal **OR** is the focus of this thesis. The work surrounding the development of this estimator and its considerations are discussed later in this chapter.

This chapter lays out the conceptual background to the novel extensions of the mediation formula, the adaptation of the extension for a simulation-based estimator and the construction of a causal **OR** from the simulation-based estimator. Each of these will be discussed in detail in the subsequent sections.

4.2 Estimation methods: A tale of two types of estimators

There are two classes of methods for estimating mediated causal effects which I focused on in this thesis which are referred to as the *model-* and *simulation-based* methods. The *model-based* methods are characterised by the use of estimated model parameters and putting them within a formula to obtain an estimate of the mediated causal effects. The formulae are derived through an application of the mediation formula as discussed previously. These formulae allow one to easily obtain estimates of the mediated causal effects and will be discussed first.

The use of formulae for the estimation of the mediated causal effects dates back to when Baron and Kenny (1986) proposed the *product of coefficients* and *difference in coefficients* methods to estimate mediated causal effects. The mediation formula had not been developed then and Baron and Kenny (1986) relied on a different set of reasoning to support the formulae proposed for the estimation. The reasoning used had a significant problem in that there were no clear distinctions between the definition of the causal effect and the methods of estimation. In the absence of this distinction, the definition is thus intertwined with how the effect is estimated and it becomes difficult to discern what a causal effect actually mean. This posed a problem subsequently when extensions were developed to adapt these formulae for binary outcomes modelled using the logistic regression. An example of a problem arising from this lack of distinction was that it was unclear if the *product of coefficients* should take place before or after transformation by the logit link function of the logistic regression (MacKinnon, 2008). Subsequent work on estimating mediated causal effects relied upon the *mediation formula* (Pearl, 2001) which crucially distinguishes between definitions of causal effects and they are estimated. The definitions of causal effects in the *mediation formula* uses the **PO** framework and provides a way to express the causal effects. The estimation of causal effects then begin by evaluating each term within the expression of a causal effect defined using the mediation formula. This evaluation can take place using any valid chosen method of estimation. The mediation formula thus establishes a direct link between the definitions of mediated causal effects and the method by which they can be estimated. The method of estimation is not restricted to parametric methods although parametric methods are the focus of this thesis. Another important aspect of the *mediation formula* is that estimates derived from it using parametric methods are the same as the *product of coefficient* formulae for scenarios where the application of the *product of coefficients* is valid and

the same parametric models are used. The *product of coefficients* method can thus be thought of as drawing upon the mediation formula as it theoretical background although acknowledgedly the *product of coefficients* method precedes the mediation formula by decades. In a similar vein, the *mediation formula* also provided a way to express mediated causal effects for multiple sequential mediators, causal **OR** and, causal **OR** for multiple sequential mediators using the **PO**s framework. Importantly, these expressions derived from the mediation formula are not tied to any specific form of estimation. The following sub-section discusses this novel extension of the mediation formula.

The other class of estimation methods is the *simulation-based* methods. These are methods that rely on simulations as a primary mean to obtain estimates of the mediated causal effects. These methods are useful particularly when the expressions obtained from the mediation formula cannot be estimated simply. One such instance is when a continuous mediator is nested within a binary outcome and they are each modelled using the normal and logistic generalised linear model (**GLM**) respectively. In such a scenario, the estimate for each **PO**, when expanded and simplified, contains an integral which does not have a closed form solution. This was discussed in the previous chapter (eq. (3.5.16)). The lack of a closed form solution meant that the integral had to be estimated using other means and one such way is through simulation. An application of simulation to estimate mediated causal effects was developed and implemented in the *mediation* *R* package (Imai et al., 2010). *mediation* was built on quasi-Bayesian simulation methods first conceptualised by King et al. (2000). The simulation estimators in *mediation* are parametric estimators which makes use of the presumed distributions of the outcomes, mediators and/or parameter estimates to simulate the various **PO**s required to estimate each of the causal effect of interest. In later sections, I will discuss the novel extensions of methods of estimation in *mediation* for estimating mediated causal effects where a binary mediator is nested within a second continuous mediator or outcome. Challenges relating to the unique nature of the **OR** and important modifications to *mediation* were also discussed. These modifications focused on correcting the errors and improving the efficiency of the programme of *mediation*.

Put together, the model-based estimators allows the estimation of some combinations of mediators and outcomes but cannot at the moment estimate mediated causal effects for specific scenarios. The simulation-based estimators on the other hand has no limitation in what mediated causal effects it can estimate. Both forms of estimation methods rely upon the mediation formula.

This thesis extends the model-based estimation to obtain expressions for multiple sequential mediation and expressions for the estimation of causal **OR**. Building on the extensions of the model-based estimators, the simulation-based estimator was extended to estimate causal effects particularly for the scenarios for which there were no previous model-based solutions. The cumulation of the work in this thesis is in an *R* programme named *seq-med* which implements all the extensions discussed. The source code of this programme can be found in chapter G. The programme provides for the estimation of any combinations of mediators and outcomes, , including those with no previous closed-form solutions. The programme also has the ability to estimate and investigate specific hypothesis in multiple sequential mediation. Heuristics had also been built into the programme that allows it to estimate mediated causal effects for any number of sequential mediators although the correctness of the programme had only been verified for up to the two mediator case. Lastly, the programme also allows the conduct of sensitivity analyses of the estimates to test their robustness against violation of certain assumptions. This latter part on sensitivity analyses and the novel contributions of this thesis is discussed at length in chapter 6. Next I shall discuss the different model combinations and which do not have an existing solution to conduct estimation of mediated causal effects.

Table 4.1 presents the different model combinations possible with a single as well as two mediators and an outcome. Combinations here refer to the type of dependent variables, either continuous or binary. Model- and simulation based estimators were defined previously. This table assumes that binary and continuous dependent variables are modelled using the logistic and normal **GLM** respectively. Had the binary dependent variables been modelled using a probit **GLM**, there is indeed a closed form solution. This closed form solution can estimate a causal **OR** but this causal ratio has a very different interpretation from the causal **OR** obtained from a logistic regression. Given the focus on the **OR** derived from the logistic regression, solutions using the probit regression will not be considered in this thesis. Also note that some of the combinations presented had solutions using both model and simulation-based estimators. These were used to validate the newly developed estimators which is discussed later in the chapter. I will next discuss in detail how the model-estimator can be extended to accommodate two mediators.

Table 4.1: Combinations of models and availability of causal effect estimator

Y	M_2	M_1	Estimators available
continuous	nil	continuous	model and simulation-based
continuous	nil	binary	model and simulation-based
binary	nil	continuous	simulation-based only
binary	nil	binary	model and simulation-based
continuous	continuous	continuous	model and simulation-based
continuous	continuous	binary	model and simulation-based
continuous	binary	continuous	simulation-based only
continuous	binary	binary	model and simulation-based
binary	continuous	continuous	simulation-based only
binary	continuous	binary	simulation-based only
binary	binary	continuous	simulation-based only
binary	binary	binary	model and simulation-based

4.3 Model-based estimators

Modern iterations of the model-based estimators mainly rests upon the mediation formula for its conceptual underpinnings. The difference between the methods based upon the mediation formula and earlier efforts such as the *product of coefficients* methods is that the earlier methods do not have a definition of the causal effects that is independent of its estimation methods. This lack of an independent definition made it ambiguous how mediated causal effects can be estimated in different scenarios and most notably for the case of the logistic **GLM**. This clear relationship established by the mediation formula between the all the variables, independent and dependent, with the causal effect, aids not just in formulating a model-based estimator. It also informs how a simulation-based estimator can be constructed. This is due largely to the non-parametric conception of the mediation formula from which one can apply it to parametric models in a number of ways where the model-based and the simulation-based estimators are two examples.

4.3.1 Applying the mediation formula for a single mediator

I will next use the case of a single binary mediator, M, with a continuous outcome, Y, to illustrate how the mediation formula can be applied to derive an expression for the estimate for any causal effect, both mediated and direct. The use of a binary mediator with a continuous outcome as an example is because this form of nesting of **PO**, a binary mediator within another continuous or binary outcome has a tractable solution because of the discrete nature of the nested term. Since the nested mediator is discrete, it is thus possible to first enumerate all possible values of M and then use those values to simulate the Y **PO** and lastly to weight each of the Y **PO** corresponding to a discrete value of M by the probability of the value of M occurring. For reference, the direct effect (**DE**) and indirect effect (**IE**) are defined as $\mathrm{E}[Y(1, M(0)) - Y(0, M(0))]$ and $\mathrm{E}[Y(1, M(1)) - Y(1, M(0))]$ respectively.

Focusing on the two causal effect definitions, it is apparent that in order to estimate the causal effects, one will first need an estimate of the **PO**s $Y(1, M(0))$, $Y(0, M(0))$ and $Y(1, M(1))$ and then their expectations. For the purposes of illustrating how the estimation will proceed, I will use the **PO** $Y(1, M(0))$. The steps for estimating the rest of the **PO** are exactly the same.

The **PO** $Y(1, M(0))$ is a nested **PO** and by that I mean that there is a dependent variable, M with its own level of R nested within it. The solution to an estimate of this Y **PO** comes from the law of iterated expectations from which the mediation formula is based upon. The law states that in order to evaluate such a nested expression, we have to take the weighted sum of Y under $R = 1$ for all the possible values of $M(0)$. In this case, since M is binary, the possible values are 0 and 1. The weights to weigh each possible value of $M(0)$ comes from the respective probability of the $M(0)$ occurring for each possible value. This will be $\mathrm{P}(M(0) = 1)$ and $\mathrm{P}(M(0) = 0)$. An expression to estimate this causal effect is:

$$Y(R = 1, M(0) = 1) \cdot \Pr(M(0) = 1) +$$
$$Y(R = 1, M(0) = 0) \cdot \Pr(M(0) = 0)$$

So the steps in order of estimating this nested **PO** will be as follows:

1. Estimate the **GLM** model for Y and M.
2. The next three steps have to be repeated for each unique subject in the dataset and at the

end of which we will have a vector of estimates corresponding to each subject for this **PO**.

(a) Estimate $\Pr(M(0) = 1)$ from the model estimated in the first step.

(b) Estimate the following **PO** of M:

$$\Pr(M(0) = 0) \cdot (\Pr(M(0) = 0) = 1 - \Pr(M(0) = 1))$$

(c) Estimate the **PO** of interest using the expression:

$$Y(R = 1, M(0) = 1) \cdot \Pr(M(0) = 1) + Y(R = 1, M(0) = 0) \cdot \Pr(M(0) = 0)$$

3. Take the expectation over this vector for an estimate of the expectation of the **PO**, i.e. $\mathrm{E}(Y(1, M(0)))$.

It should be noted that although the steps were repeated for each subject within the dataset, this should not be interpreted as an individual causal effect. This effect is the inferred effect of an individual who possessed the same observed characteristics as the subject in the dataset. By repeating the steps for each subject in the dataset, we are using the distribution of the characteristics of the subjects to weight each of the individually estimated **PO** such that the expectation of the **PO** is really a weighted sum of the **PO** using the probability of someone possessing the same characteristic as the subject. If all the subjects were unique (as in the case where there are continuous covariates), then the weights will simply be $\frac{1}{n}$ where n is the sample size.

Next, we need to repeat these steps for each **PO** that is required. This is informed by the causal effects definition. The final step once the expectations of the **PO** had been estimated is to plug in the value of each **PO** into the causal effect definition and compute the causal effect estimate according to the definition. In the context of this thesis, if the outcome was continuous, we would take the difference between the expectations of the two **PO**. If it was binary, we would use the mean predicted probabilities of each **PO** and compute a causal **OR**.

4.3.2 Applying the mediation formula for two mediators

In order to extend the model-based estimator to accommodate two sequential mediators, we will follow the same steps as the single mediator scenario with a few exceptions. First, there

are more mediators and since we are interested in all possible mediated pathways, the second mediator, M_2, will have the first mediator, M_1, nested within it. The outcome, Y, will have both the M_1 and M_2 nested within it. This results in the **PO** of Y having two levels of nesting as M_2 is also nested compared with a single level of nesting when there was only a single mediator.

The causal effects of interest are as follows:

$$\mathbf{DE} = \mathrm{E}[Y(1, M_1(0), M_2(0, M_1(0)))] - \mathrm{E}[Y(0, M_1(0), M_2(0, M_1(0)))]$$

$$\mathbf{IE}_1 = \mathrm{E}[Y(1, M_1(1), M_2(0, M_1(0)))] - \mathrm{E}[Y(1, M_1(0), M_2(0, M_1(0)))]$$

$$\mathbf{IE}_2 = \mathrm{E}[Y(1, M_1(1), M_2(1, M_1(0)))] - \mathrm{E}[Y(1, M_1(1), M_2(0, M_1(0)))]$$

$$\mathbf{IE}_3 = \mathrm{E}[Y(1, M_1(1), M_2(1, M_1(1)))] - \mathrm{E}[Y(1, M_1(1), M_2(1, M_1(0)))]$$

Similar to the single mediator scenario, each causal effect had been defined as a difference between two **PO**s since Y is continuous. Similarly, since the methods used for estimating a single **PO** can be applied to any **PO**, I will use $\mathrm{E}[Y(1, M_1(1), M_2(0, M_1(0)))]$ to illustrate how the estimation proceeds.

There are two important implications of having another level of nesting. First, since M_1 is nested within M_2, evaluating M_2 will require the use of the same methods we used for evaluating the outcome in the single mediator scenario. In this case, instead of the outcome, we have M_2 and similarly we have a single mediator. We can adopt the expression derived from the single mediator scenario to estimate the **PO**s of M_2. Consider the scenario where we have only a single mediator and an outcome. This scenario was discussed previously and the same methods used to estimate such a **PO** can be used to estimate the **PO** when a mediator is nested within another mediator. Secondly, since there are two mediators, summing across all 'possible values of the mediator' previously now becomes summing across all 'possible combination of values of the mediators'. I will use the scenario of two binary sequential mediators and a continuous outcome to illustrate how the mediation formula can be extended. For two binary mediators, we will have four possible combinations, i.e. $M_1 = 0$ with $M_2 = 0$, $M_1 = 0$ with $M_2 = 1$, $M_1 = 1$ with $M_2 = 0$, and $M_1 = 1$ with $M_2 = 1$. Each of these combinations have their respective probability of occurring which will need to be estimated and used as weights when summing the **PO** of Y under each of these combinations, similarly to what we did for a single mediator. Also as before,

once all the **PO** had been estimated, we can then take their expectations and compute the causal effect according to the causal effect definition.

The simulation of each pair of **PO**s for the estimation of a causal effect follows closely the way the mediation formula is structured. This coupling of the estimation method with a theoretical base provides a strong intuition on how and why the estimation procedure works. It further provides clues as to how this estimation procedure can be modified to address similar problems. I will next discuss how this procedure for estimating **PO** with a single mediator is modified to estimate **PO** with two mediators.

Returning to the estimation of the **PO**, $\mathrm{E}[Y(1, M_1(1), M_2(0, M_1(0)))]$. The first task is to work out the order in which the **PO**s of Y had to be estimated. Y **PO**s can not be estimated without first estimating M_1 and M_2 **PO**s. M_2 **PO**s can not first be estimated without estimating M_1 **PO**s. Therefore, M_1 **PO**s had to be estimated first because it does not have a dependency on other **PO**s. Once M_1 **PO**s had been estimated, we can then estimate M_2 **PO**s but not Y **PO**s. Only when M_1 and M_2 **PO**s had been estimated can we estimate Y **PO**s. The steps to conduct the estimation are as follows:

1. Estimate the **GLM** models for Y, M_1 and M_2.
2. Determine the order in which the **PO**s have to be estimated and as discussed, this will be M_1, M_2 then Y for the two mediators scenario.
3. Start by estimating the M_1 **PO** required. This will be $M_1(0)$ and $M_1(1)$ for the example of Y **PO** which we are using as an example. This estimation uses the model of M_1 estimated in (1) and estimate M_1 for each subject present in the dataset while setting the level of R to 0 and 1 for the respective M_1 **PO**s.
4. The next step is to estimate **PO** of M_2, $M_2(0, M_1(0))$. Since the **PO** contains M_1 **PO** nested within it, we will need to apply the mediation formula to estimate this **PO**. Following the steps outlined in the single mediator scenario when evaluating the **PO** of Y, we will end up with the expression, $M_2(0, M_1(0) = 1) \cdot \Pr(M_1(0) = 1) + M_2(0, M_1(0) = 0) \cdot \Pr(M_1(0) = 0)$. This expression allow us to estimate the M_2 **PO**, $M_2(0, M_1(0))$.
5. Having obtained an expression to estimate M_2, we will next need to consider the possible combinations of M_1 and M_2. Since each of them are binary, we will have the following combinations:

- $M_1(0) = 0, M_2(0, M_1(0)) = 0$
- $M_1(0) = 0, M_2(0, M_1(0)) = 1$
- $M_1(0) = 1, M_2(0, M_1(0)) = 0$
- $M_1(0) = 1, M_2(0, M_1(0)) = 1$

6. We will next need to obtain the probability of each of these combinations occurring. This can be obtained by taking the product of probability of M_1 and M_2 occurring. This is a valid estimation of the probability of the combination occurring because the probability of M_2 occurring is a conditional probability, conditional on both M_1 and R and any other covariates that might be present.
7. Once the probabilities of each combination had been obtained, we then use it to weight each of the Y **PO** under each combination which will yield the following expressions.
   - $\Pr(M_1(0) = 0) \cdot \Pr(M_2(0, M_1(0)) = 0) \cdot Y(1, M_1(0) = 0, M_2(0, M_1(0)) = 0)$
   - $\Pr(M_1(0) = 0) \cdot \Pr(M_2(0, M_1(0)) = 1) \cdot Y(1, M_1(0) = 0, M_2(0, M_1(0)) = 1)$
   - $\Pr(M_1(0) = 1) \cdot \Pr(M_2(0, M_1(0)) = 0) \cdot Y(1, M_1(0) = 1, M_2(0, M_1(0)) = 0)$
   - $\Pr(M_1(0) = 1) \cdot \Pr(M_2(0, M_1(0)) = 1) \cdot Y(1, M_1(0) = 1, M_2(0, M_1(0)) = 1)$
8. The last step to obtain an estimate of the Y **PO** of interest is simply to sum up the estimates from the previous step.
9. To obtain an expectation of the Y **PO** of interest, we average across all the estimated Y **PO** for each subject.
10. We will repeat steps (1) - (9) for each **PO** of interest.
11. Once the expectation of all the **PO**s required had been estimated, we will then use the estimated expectations according to the causal effect definition to produce an estimate of the causal effect.

The steps presented illustrate how the mediation formula had been extended for two mediators. Conceptually, the same method can be used to extended to estimate the mediated causal effects of any number of mediators. This can be done by iteratively applying the steps presented, starting with the mediator with no nested **PO** to the mediator with the most number of nested **PO**s. The programme for estimating the mediated causal effects has this iterative method implemented and can theoretically be used to estimate causal effects of any number of mediators but the performance of the programme had only been validated up to the two mediators case. Also, it needs to be clarified that while the mediated causal effects of any number of mediators can be estimated, the necessity and motivation of which will need to be justified appropriately.

So far we have focused the discussion to binary mediators and continuous outcomes. If the mediators are continuous, instead of using the probability of the mediator occurring, which is derived from the probability mass function (**PMF**) of the mediator, we will use the integral of the probability density function (**PDF**) of the mediator instead. If the outcome is binary, our causal effect will be defined as **OR** instead of differences. A problem arises however when the mediators are continuous and the outcome is binary. Since the integral of the **PDF** of the mediator is used instead of the **PMF**, this will give rise to the integral that does not have a closed form solution (eq. (3.5.16)) discussed in the previous chapter. These scenarios also occur when the first mediator is continuous and the second mediator is binary since the evaluation of the **PO** of the second mediator will also require the evaluation of the integral. The next section discusses the development of a novel estimator to estimate the causal effects in such scenarios. This novel estimator builds upon the model-based estimator and provides a way to estimate the causal effects using simulation to estimate the causal effects where the model-based estimator does not have a closed form solution. This novel estimator is referred to as the *simulation-based estimator*.

In summary, this section first discussed how the model-based estimator, built upon the mediation formula, is applied to estimate mediated causal effects for a single mediator and then subsequently extended for two mediators. The next section builds upon the model-based formula and its extensions to create a novel simulation-based estimator which allows for the estimation of any combination of mediator and outcomes, including the ones where the model-based estimator does not have a closed form for.

4.4 Simulation-based estimators

The simulation method used in this thesis is a quasi-Bayesian method developed by King et al. (2000) originally for the purpose of policy analysis. The authors were interested in answering questions such as 'what will the outcome be if we altered the policy in a certain manner?' by being able to simulate outcomes had the policy been changed in various way. Such questions are in fact questions about comparing different **PO**s with each outcome being the result had we altered the policy one way or another. These methods were thus well suited for simulation of **PO**s as a means to estimate mediated causal effects. Notably, Imai et al. (2010) used these

methods for an estimator of mediated causal effects in the *R* package *mediation*. The work presented in this thesis follows closely the implementation used by Imai et al. (2010).

The main idea behind this quasi-Bayesian simulation is the idea that there are uncertainties associated with estimation of a parametric model such as the fitting of a **GLM**. The authors identified two main forms of uncertainties and named them *estimation* and the *fundamental uncertainty*. Since the parameters are each estimated with uncertainty, it necessarily implies that there is a range of values which the estimated parameter can lie within. This range of values and the associated probabilities of each value occurring thus forms a distribution of the estimated parameter and put in another way, the point estimate of the parameter is drawn from this distribution of values. The idea behind the simulation is that if we repeatedly make draws of the parameters from this distribution and using these draws to make an estimate of the outcome, we will be able to form a distribution of the outcome which factors in the uncertainties. Going one step further, if we make these repeated draws while keeping an independent variable fixed at a value of our choosing such as the treatment level in an randomised controlled trial (**RCT**), we will then be able to simulate **PO**s.

The 'Bayesian' part of this simulation procedure comes from the use of the *prior* distribution(s) of the parameters and the observed data from the sample to form a posterior distribution of the outcome. The *prior* distribution is not a true prior distribution in the Bayesian sense since it is a distribution formed from the initial estimation of the model and hence dependent on the sample data. This is the reason this simulation method is termed a quasi-Bayesian rather than a Bayesian method. I will next discuss the two forms of uncertainty, how they are quantified and how they can be used to form the prior distribution of parameters from which to draw from.

4.4.1 Estimation uncertainty

Estimation uncertainty, as the name suggest is the uncertainty that arise from the estimation process. This can be due to various reasons including but not restricted to a small sample set and the choice of estimation methods. In the context of using a **GLM** to model the outcome of interest, the regression coefficients are parameters that are estimated to quantify the relationship between the outcome and the independent variables. These coefficients, being estimated quantities, are

measured with a level of uncertainty that is quantified by the variance-covariance matrix of the coefficients. This matrix is a symmetrical matrix with the number of rows being the number of independent variables. The diagonal is the variance of the corresponding regression coefficient while the rest of the matrix contains the pairwise covariance between the regression coefficients. The confidence interval of the regression coefficients commonly presented in studies is derived from this matrix.

One way to simulate the regression coefficient is to make use of this matrix. What the variance-covariance matrix conveys is not just the variance of each individual coefficient but also their relationship in the form of a covariance with the other coefficients. If we make the assumption that the underlying distribution of the regression coefficients is a multivariate normal distribution (**MVN**) distribution, we can parametrise a **MVN** distribution using the regression coefficients as means and the matrix as the variance-covariance matrix. We can then treat this distribution as the underlying distribution of the regression coefficient from which we can then make draws from to simulate the regression coefficients. An alternative was if we do not wish to make the **MVN** distribution assumption will be to use a non-parametric bootstrap. Generally, this is done by first sampling the sample with replacement and then fit the model again. The regression coefficient from this second fitting is then treated as a draw from the underlying distribution of the regression coefficients. The assumption of **MVN** is justified for large sample sizes since maximum likelihood was used to fit the model. This provides for a statistically more powerful approach and is also computationally less intensive then the bootstrap approach. The bootstrap approach should be used when the sample size is small. Both methods had been proposed by King et al. (2000) as valid ways to simulate the coefficients and have been implemented in the *mediation* package by Imai et al. (2010) as well as in the implementation used in this thesis.

4.4.2 Fundamental uncertainty

The other form of uncertainty is the fundamental uncertainty. This is the uncertainty associated with random events or unknown causes that have the ability to influence the outcome. An example of such uncertainty can be illustrated using a coin toss. For a perfectly balanced coin, the chance of the coin landing on heads or tails is each 50%. Knowing this probability only tells us that for a large number or infinite number of coin tosses, the coin will land on heads 50% of

the time but it does not allow us to say with certainty which side the coin will land for the next toss.

In the context of a normal **GLM**, this is quantified by the residual variance. This variance is the variance that is not explained by all the other independent variables and hence represent the variance of the outcome for which we do not know the causes of. The normal **GLM** parametrises the residuals to have a mean of 0 and together with the estimate of the residual variance, we can parametrise a normal distribution with a mean of 0 and variance using the estimated variance. This distribution can then be used to make draws from to simulate the residual variance. For logistic regressions, the process is a little different which will be discussed in section 4.5.1.

4.4.3 Simulating outcomes

Having discussed the two forms of uncertainties and the general way in which they can be used to simulate the regression coefficients and residuals, I will next describe the steps that are used to conduct the simulation of a simple **PO** where we are interested in the outcome had the binary treatment indicator R been held at level 1. This outcome is written as $Y(1)$. The following discussion will use many of the notations set out at the beginning of the thesis and should be used as a reference.

For this example, I will be using a simple normal **GLM** with the identify link function f where:

$$Y = f(R, C) = \beta_0 + \beta_1 R + \beta_2 C + \varepsilon,\ \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

where Y is the outcome of interest (continuous).

R is the treatment level (binary).

C is the covariate (can be either binary or continuous).

ε is the residual.

The simulation will proceed as follows:

1. Fit model Y as specified.
2. Extract the vector of estimated regression coefficients, $\vec{\boldsymbol{\beta}}$.
3. Extract the variance-covariance matrix of $\vec{\boldsymbol{\beta}}$, $cov(\vec{\boldsymbol{\beta}})$.
4. Extract the **MSE** of the model.

5. Set the value of R to 1.
6. Simulation steps:
   (a) Simulate the regression coefficients of model Y by making a draw from the distribution $\mathcal{N}_{\text{MV}}(\vec{\boldsymbol{\beta}}, cov(\vec{\boldsymbol{\beta}}))$. The draw, $(\vec{\boldsymbol{a}})_k$, is a vector of length p and the k subscript represents a single run of the simulation. It represents a single draw of the regression coefficients from their assumed underlying distribution.

$$(\vec{\boldsymbol{a}})_k = (\{a_1, a_2, \ldots, a_p\} \sim \mathcal{N}_{\text{MV}}(\vec{\boldsymbol{\beta}}, cov(\vec{\boldsymbol{\beta}})))_k$$

   (b) Simulate the estimated, expected outcomes for each subject, $(\vec{\boldsymbol{b}})_k$ by making a 'prediction' of Y using $(\vec{\boldsymbol{a}})_k$ and the model matrix $\boldsymbol{X}_Y$ (sample data used in the estimation of the Y model) by taking the dot product of the matrix with the vector of simulated regression coefficients. The resulting vector is assigned to $(\vec{\boldsymbol{b}})_k$. $(\vec{\boldsymbol{b}})_k$ is a vector of simulated Y taking into account the estimation uncertainty.

$$(\vec{\boldsymbol{b}})_k = (\boldsymbol{X}_Y \times (\vec{\boldsymbol{a}})_k)_k$$

   (c) Simulate the fundamental uncertainty, $(\vec{\boldsymbol{d}})_k$ by making n draws from the distribution $\mathcal{N}(0, \text{MSE})$.

$$(\vec{\boldsymbol{d}})_k = (\{d_1, d_2, \ldots, d_n\} \sim \mathcal{N}(0, \text{MSE}))_k$$

   (d) Add the fundamental uncertainty to $(\vec{\boldsymbol{b}})_k$ (The resulting vector $(\vec{\boldsymbol{g}})_k$ is the simulated Y for the kth run of the simulation which takes into account both forms of uncertainty).

$$(\vec{\boldsymbol{g}})_k = (\vec{\boldsymbol{b}})_k + (\vec{\boldsymbol{d}})_k$$

   (e) Summarise the results of the kth simulation by taking the expectation of $(\vec{\boldsymbol{g}})_k$, $\text{E}[(\vec{\boldsymbol{g}})_k]$
7. Repeat step 6 for m times. At the end of the simulations, we will get a vector, $\vec{\boldsymbol{h}}$, of length m. $\vec{\boldsymbol{h}}$ is the vector containing the expectations of simulated Y from each individual simulation run.

$$\vec{\boldsymbol{h}} = \{\text{E}[(\vec{\boldsymbol{h}})_1], \text{E}[(\vec{\boldsymbol{h}})_2], \ldots, \text{E}[(\vec{\boldsymbol{h}})_m]\}$$

8. A point estimate for the **PO**, $Y(1)$ can then be obtained by taking the expectation of $\vec{\boldsymbol{h}}$ to get $\text{E}[\vec{\boldsymbol{h}}]$.

I will discuss how this can be applied to estimate mediated causal effects in the next section.

4.5 Estimating mediated causal effects by simulation

I will use an example of a single mediator to illustrate how this simulation procedure can be used to estimate a mediated causal effect. For this example, I will use the case of a continuous mediator and outcome. Although this combination of mediator and outcome have existing model-based solutions, this illustration is meant to provide a reference method by which the mediated causal effect can be estimated by simulation. Subsequent discussion will touch on the modifications to this method to allow for the estimation of mediated causal effects involving a mixture of binary and continuous variables.

Assuming a continuous mediator and outcome, we have the following models:

$$Y = f(R_Y, C, M) = \beta_0 + \beta_1 R_Y + \beta_2 C + \beta_3 M + \varepsilon_Y, \quad \varepsilon_Y \sim \mathcal{N}(0, \sigma_Y^2)$$
$$M = f(R_M, C) \quad = \alpha_0 + \alpha_1 R_M + \alpha_2 C + \varepsilon_M, \qquad \varepsilon_M \sim \mathcal{N}(0, \sigma_M^2)$$

where Y is the outcome of interest (continuous)

M is the mediator (continuous)

R is the treatment level (binary)

C is the covariate (can be either binary or continuous)

ε_Y and ε_M are residuals of the respective models.

For this example, I will be simulating the expectation of the **PO** $Y(1, M(0))$. The steps to simulate the mediated causal effects are as follows:

1. Fit the models, Y and M as specified.
2. Extract parameters from the Y model.
   (a) Extract the vector of estimated regression coefficients of Y, $\vec{\boldsymbol{\beta}}$.
   (b) Extract the variance-covariance matrix of $\vec{\boldsymbol{\beta}}$, $cov(\vec{\boldsymbol{\beta}})$.
   (c) Extract the **MSE** of Y, $\mathbf{MSE}_Y$.
3. Extract parameters from the M model.
   (a) Extract the vector of estimated regression coefficients of M, $\vec{\boldsymbol{\alpha}}$.
   (b) Extract the variance-covariance matrix of $\vec{\boldsymbol{\alpha}}$, $cov(\vec{\boldsymbol{\alpha}})$.
   (c) Extract the **MSE** of M, $\mathbf{MSE}_M$.
4. Set the values of R to the required levels for the **PO** $Y(1, M(0))$.

(a) Set the value of R_Y to 1.

(b) Set the value of R_M to 0.

5. Simulation steps:

   (a) Simulate M.

      i. Simulate the regression coefficients of model M, $(\vec{\boldsymbol{a}})_k$ by making a draw from the distribution $\mathcal{N}_{\text{MV}}(\vec{\boldsymbol{\alpha}}, cov(\vec{\boldsymbol{\alpha}}))$ and assign the resulting vector to $(\vec{\boldsymbol{a}})_k$.

      ii. Simulate the estimated, expected M for each subject, $(\vec{\boldsymbol{b}})_k$ by making a 'prediction' of M using $(\vec{\boldsymbol{a}})_k$ and the sample data $\boldsymbol{X_M}$ to get $(\vec{\boldsymbol{b}})_k$.

      iii. Simulate the fundamental uncertainty of M, $(\vec{\boldsymbol{d}})_k$ by making a draw from the distribution $\mathcal{N}(0, \mathbf{MSE}_M)$.

      iv. Add the fundamental uncertainty to the estimated expected M for each subject to get $(\vec{\boldsymbol{d}} + \vec{\boldsymbol{b}})_k$ ($(\vec{\boldsymbol{d}} + \vec{\boldsymbol{b}})_k$ is the simulated M for the M **PO** $M(0)$ for the kth run of the simulation).

   (b) With M simulated, we will next use it to simulate Y.

      i. Simulate the regression coefficients of model Y, $(\vec{\boldsymbol{g}})_k$ by making a draw from the distribution $\mathcal{N}_{\text{MV}}(\vec{\boldsymbol{\beta}}, cov(\vec{\boldsymbol{\beta}}))$ and assign the vector to $(\vec{\boldsymbol{g}})_k$.

      ii. Set the value of M to $(\vec{\boldsymbol{d}} + \vec{\boldsymbol{b}})_k$ in the model matrix $\boldsymbol{X_Y}$.

      iii. Simulate the estimated, expected Y for each subject by making a 'prediction' of Y using $(\vec{\boldsymbol{g}})_k$ and the modified model matrix $\boldsymbol{X_Y}$ to get $(\vec{\boldsymbol{h}})_k$.

      iv. Simulate the fundamental uncertainty of Y, $(\vec{\boldsymbol{q}})_k$ by making a draw from the distribution $\mathcal{N}(0, \mathbf{MSE}_M)$ and assign it to $(\vec{\boldsymbol{q}})_k$.

      v. Add the fundamental uncertainty ($(\vec{\boldsymbol{q}})_k$) to the estimated expected Y ($(\vec{\boldsymbol{h}})_k$) and assign the result to $(\vec{\boldsymbol{t}})_k$.

      vi. Take the expectation of $(\vec{\boldsymbol{s}})_k$, $\text{E}[(\vec{\boldsymbol{s}})_k]$ ($\text{E}[(\vec{\boldsymbol{s}})_k]$ represents a single draw of Y from its underlying distribution accounting for both the estimation and fundamental uncertainties).

6. Repeat step 6 for m times, the total number of simulations required. Once the number of simulations is reached, we will have a set of $\text{E}[(\vec{\boldsymbol{s}})_k]$, one each from each simulation. We assign this vector $\vec{\boldsymbol{t}}$.

7. The point estimate of the **PO** $Y(1, M(0))$ can be obtained from $\vec{\boldsymbol{t}}$ by taking its expectation, $\text{E}[\vec{\boldsymbol{t}}]$.

The steps outlined allow us to simulate and obtain estimates for each **PO**. These then have to be

used according to our causal effect definitions in order to obtain an estimate of the causal effect. If we are interested in the direct effect, defined as $\mathrm{E}[Y(1, M(0))]$ - $\mathrm{E}[Y(0, M(0))]$, then we will do the following:

1. For the kth run of the simulation, do the following.
   (a) Simulate the **PO** $M(0)$.
   (b) Simulate the **PO** $Y(1, M(0))$ using the $M(0)$ simulated in the first step.
   (c) Simulate the **PO** $Y(0, M(0))$ using the $M(0)$ simulated in the first step.
2. At the end of the simulations, we will take the expectation of each of the simulated **PO** before taking the difference between the expectation of the **PO**s. This provides us with a point estimate of the direct causal effect.
3. The confidence interval can be constructed by using the two vectors of simulated **PO**, taking their differences and using the 97.5th and 2.5th percentile values from the resulting vector as the upper and lower bounds respectively. This is known as the percentile method.

Note the two **PO** in question share a common, simulated term, $M(0)$. The same $M(0)$ needs to be used when simulating the Y **PO**. This is a reflection of the definition of **PO** which states that all conditions are to remain the same except for a single R. The R that is different indicates the effect that we are interested in and in this example, the R that is different is R_Y.

Next I will discuss some variations to this simulation starting with the extension to accommodate logistic regression. The method of simulation just outlined will be referred to as the *standard method of simulation.* Other methods will bear reference to this method.

4.5.1 Variation I: Logistic regression, fundamental uncertainty

Logistic **GLM** is used to model binary outcomes and mediators in this thesis. While the normal **GLM** has a residual of mean 0 and unknown variance which is to be estimated, the logistic **GLM** has a residual with a fixed mean and variance to that of a standard logistic distribution. The residual of a logistic regression is fixed in order to identify the model which then allows the estimation of the parameters of interest to move forward. This requirement exist because the logistic regression is modelling a latent attribute of the outcome, its underlying continuous distribution, rather than the outcome itself.

The implication of this to the study is that the use of the estimated variance of the residual to simulate the fundamental uncertainty in the normal **GLM**, is no longer usable in the logistic instance. Here I will discuss an alternative way to simulate this fundamental uncertainty and the considerations of such an alternative. Imai et al. (2010) in the *mediation* package, simulated the fundamental uncertainty as follows:

1. For a given run of the simulation, after computing the predicted probability for each subject present in the sample, the probability is then used to simulate the fundamental uncertainty.
2. For an outcome Y, assuming that the predicted probability for outcome Y is 0.8, I will then make a 'coin toss' of a loaded coin with a probability of landing on head 80% of the time and record the result. The coin toss is actually implemented as a simulation procedure in the programme.
3. The result is then used as the simulated outcome. This result incorporates both the fundamental and estimation uncertainty in its simulation because the regression coefficients were drawn from a distribution of the coefficients parametrised by the estimation uncertainty and the fundamental uncertainty is embedded in the predicted probability itself.

While this procedure does simulate the fundamental uncertainty, the uncertainty itself is not reproducible for use in the simulation of other **PO**s. If each **PO** that needs to be simulated were to be simulated in this manner, they will all have a different magnitude and direction of the fundamental uncertainty. Recall that in the standard method, the fundamental uncertainty is simulated once and shared across **PO**s for a given subject within a given simulation run. This sharing of the fundamental uncertainty arises from the definition of our causal effects. The causal effects are defined by comparing two **PO**s either using differences for continuous outcomes or **OR** for binary outcome. The two **PO**s are measured under *identical* conditions but for the level of a single R. The level of R chosen depends on which effect is of interest and since two **PO**s are measured under the same conditions, changes in the outcome can thus be attributable to the change in the level of R. If the two **PO**s had different fundamental uncertainty components, the difference between outcomes can also be due to differences in the fundamental uncertainty and is not solely attributable to R.

In order to use the same fundamental uncertainty component for both **PO**s, we need to develop a mechanism to quantify and simulate the fundamental uncertainty and use this simulated quantity for both **PO**s, similar to how the fundamental uncertainty is simulated in the standard approach. For this purpose, I turn to the latent variable formulation of the logistic regression discussed earlier (eqs. (3.5.3) and (3.5.4)). For reference, the equations are reproduced here:

$$\begin{aligned} Y &= \mathrm{I}((Y^* + \varepsilon) \geqslant 0) \\ Y^* &= \mathrm{logit}(\Pr(Y = 1)) \\ &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \end{aligned}$$

As a recap, Y^* is a latent variable which is conceptualised as the logit transform of $\Pr(Y = 1)$. 'I' is the indicator function that returns 0 if its argument evaluates to *false* and 1 otherwise. Y is predicted by adding a random number drawn from $\mathcal{L}_{\mathbf{STD}}$(represented by ε) to the estimated Y^* and passing the result through the I function.

As we can see here, the latent variable formulation of the logistic regression has an explicit term, the ε, which represents the fundamental uncertainty. Using this formulation, we can then ensure that the **PO**s within a subject for a given simulation shares a common fundamental uncertainty component. Therefore, to simulate the fundamental uncertainty for a given subject, we will replace steps 5(a)iii, 5(a)iv, 5(b)iv and 5(b)v in the simulation of mediated outcomes depending on which of M and Y is/are binary. For each subject in a given simulation,

1. Simulate the fundamental uncertainty by making a draw from $\mathcal{L}_{\mathbf{STD}}$.
2. Add the draw to the predicted probability of the outcome.
3. If the resulting sum is greater than 0, assign 1 for the simulated outcome and 0 otherwise.

We can then proceed using the simulated Y or M as in the standard method. This method therefore allow us to simulate binary outcomes while ensuring that the fundamental uncertainty component is kept constant for a subject within a simulation and thus adhere to the definition of **PO**.

For clarity, the steps to simulate a binary Y is reproduced below, using the **PO** $Y(1)$ as an example.

1. Fit model Y^* as specified.

2. Extract the vector of estimated regression coefficients, $\vec{\boldsymbol{\beta}}$.
3. Extract the variance-covariance matrix of $\vec{\boldsymbol{\beta}}$, $cov(\vec{\boldsymbol{\beta}})$.
4. Set the value of R to 1 in the model matrix $\boldsymbol{X}_Y$
5. Simulation steps:
   (a) Simulate the regression coefficients of model Y^*, $(\vec{\boldsymbol{a}})_k$.
   Make a draw from the distribution $\mathcal{N}_{\text{MV}}(\vec{\boldsymbol{\beta}}, cov(\vec{\boldsymbol{\beta}}))$.
   (b) Simulate the estimated logit$(\Pr(Y = 1))$ for each subject, $(\vec{\boldsymbol{b}})_k$.
   Make a prediction of Y^* using $(\vec{\boldsymbol{a}})_k$ and the model matrix $\boldsymbol{X}_Y$
   (c) Make a draw from $\mathcal{L}_{\text{STD}}$for each subject and add this to $(\vec{\boldsymbol{b}})_k$ to obtain $(\vec{\boldsymbol{d}})_k$.
   (d) Summarise the results of the kth simulation by taking the expectation of $(\vec{\boldsymbol{d}})_k$, $\text{E}[(\vec{\boldsymbol{d}})_k]$
6. Summarise the results of the kth simulation by taking the expectation of $\text{I}(\text{E}[(\vec{\boldsymbol{d}})_k \geqslant 0])$ to obtain a simulated Y.
7. Repeat step 5 for m times.
   At the end of the simulations, we will get a vector, $\vec{\boldsymbol{g}}$, of length m.
   $\vec{\boldsymbol{g}}$ is a vector containing the expectations of simulated Y from each individual simulation run.
8. A point estimate of the **PO** $\Pr(Y(1) = 1)$ can be computed by taking the expectation of $\vec{\boldsymbol{g}}$, $\text{E}[\vec{\boldsymbol{g}}]$.

Using the simulation of a binary M as an example, if we were to simulate M using the above modified procedure an infinite number of times and take the expectation of the resulting simulated M, we will recover the estimated $\Pr(M = 1)$. This is to be expected since the estimated $\Pr(M = 1)$ is summary of M if we were able to somehow sample M repeatedly and infinitely from the population. This is a more efficient method of simulating binary mediators and in the next variation, I will discuss how this can be done.

4.5.2 Variation II: Logistic regression, application of the mediation formula

When working with binary outcomes and mediators, I have shown in the previous section how these binary variables can be simulated while adhering to the definition of **PO**. Here I will

present and alternative that can be used in the case of binary mediators by making use of the mediation formula.

Recall that in the model-based approach, binary mediators are used by taking a probability weighted sum of the possible values of M. In the case of simulation of binary mediators, we can apply the same concept. To illustrate this, consider a binary M and a continuous Y and the **PO** $Y(1, M(0))$.

In the model-based approach, we will estimate the $Y(1, M(0))$ using the following expression:

$$Y(1, M(0)) = Y(1, M(0) = 1) \cdot \Pr(M(0) = 1) + Y(1, M(0) = 0) \cdot \Pr(M(0) = 0)$$

Similarly, we can apply the same mediation formula within a simulation run to obtain an estimate of $Y(1, M(0))$ This allow us to make the simulation more efficient because captured within the estimate of $Y(1, M(0))$ is not just a single draw of M but incorporates the long run expectation of M. The alternative is to simulate M and realise it as a binary variable for each simulation. Doing so over a large number of simulations should yield the same results as not realising M, i.e. by applying the mediation formula and using the predicted $\Pr(M = 1)$. Both of these methods, realising and not realising a binary dependent variable, will yield the same results given a large number of simulations. The same estimate could be obtained using smaller number of simulations by not realising the dependent binary variable and therefore, binary dependent variables are not realised in the estimation of causal effects involving binary dependent variable.

The example given was for a continuous Y. What happens then if Y is binary? Since we have now established methods for simulating binary variables while maintaining the same fundamental uncertainty, there is now the option as to which method should be use when estimating the causal **OR** of a binary Y. The next section will discuss this.

4.5.3 Variation III: Logistic regression, estimating causal OR

For a binary outcome Y, there are two levels at which we have the option of realising Y. The first is within a single simulation. For each **PO** within a simulation, each subject will have a $\Pr(Y = 1)$ and at the subject level within each simulation, we can choose to either realise Y or let it remain as a probability. Regardless of whether we choose to realise Y within a simulation, at

the end of a single run of the simulation, the expectation of each **PO** is a probability of Ydrawn from the underlying distribution of Y. The second level is whether or not this probability is to be realised.

Before considering the options presented, we first need to consider the causal **OR**. **OR** differs from differences as causal effects in one important way. Differences as causal effect are collapsible while the **OR** is not as discussed previously. This means that for a continuous outcome the order in which we compute the differences and taking the expectations does not matter.

$$\mathrm{E}[Y(1)] - \mathrm{E}[Y(0)] = \mathrm{E}[Y(1) - Y(0)]$$

This relationship between expectations of differences and differences of expectations means that for a continuous outcome, it does not matter whether we take the differences at the subject level within a simulation and then the expectation of the differences or the differences between expectations of each **PO** within a simulation. This also suggests that there are certain terms that can be skipped in the simulation such as covariates since they cancel out each other as discussed previously.

The **OR** however is different and because of the lack of collapsability, the results if we were to compute the **OR** at the individual level with a simulation and taking the expectation is different from the result if we take the expectation of the simulated outcomes and using these expectations (which will be probabilities) or each **PO** to compute the **OR**. Since we also have two levels at which we can choose to realise the binary variable or not, it also implies that there are two levels where we can compute the **OR**, one at the subject level within a simulation and the second across all the simulations.

The choice for how to proceed rests upon the definition adopted for causal **OR**. Specifically, in this thesis, I am defining the causal **OR** to be the marginal **OR**. Therefore stemming from this definition of the causal effect, the marginal **OR** will thus need to be computed from marginal predicted probabilities of each **PO**. So therefore, instead of summarising each simulation at the end of it, we will hold the two **PO**s from each simulation and at the end of all the simulations, we will have two vectors of simulated **PO**. These two vectors consist of simulated probabilities of the binary outcome. We can then take the expectation for each of these vectors and obtain the simulated probability of each of the two **PO**. This simulated probability can then be used to construct a causal **OR**. Since this **OR** is constructed from estimates of the marginal probabilities

of each **PO**, this is then the marginal **OR**.

Having discussed how we can obtain a causal **OR** despite the non-collapsibility of **OR**, we revisit the question of whether or not the outcome should be realised. For both the within simulations and the across simulations level, I have opted not to realise binary outcomes for the same reason that I have chosen not to realise binary mediator, i.e the predicted probabilities is a sufficient estimate of the simulated outcome. Realising the binary outcome over a large number of simulations will yield the same estimates as not realising but will incur a large computational cost. I should note that in the programme written to carry out the simulation, the option is present whether or not to realise binary M and Y.

Now that we can estimate causal effects for all combinations of binary and continuous M and Y, we can then consider how this can be extended to cover the two mediators case.

4.5.4 Variation IV: Two mediators

The extension of the simulation estimator to accommodate two mediators follows closely what was laid out in the model-based approach as follows using the example of estimating $\mathrm{E}[Y(1, M_1(1), M_2(0, M_1(0)))]$. For this example, assume that M_1, M_2 and Y are all binary. The use of all binary dependent variables is for simplicity because the extension of the model with all binary dependent variables requires only the predicted probability of each of the dependent variable and does not require the use of any integrals of probability distribution functions as is the case with continuous dependent variables. The extension proceeds as follows:

1. Estimate models for Y, M_1 and M_2.
2. Determine the order in which the **PO**s need to be estimated and the order is M_1, M_2 and Y.
3. Identify the **PO**s required. These are

$$M_1(1),\ M_1(0),$$
$$M_2(0, M_1(0)),\ \text{and}$$
$$Y(1, M_1(1), M_2(0, M_1(0)))$$

4. Estimate $M_1(1)$.

(a) Simulate $M_1(1)$ and $M_1(0)$ using the simulation methods previously discussed.

(b) At the end of each simulation, we will have a vector of simulated predicted probabilities of M_1, one for each subject and **PO**s required.

5. Estimate $M_2(0, M_1(0))$.

(a) Identify the different combinations of $M_1(0)$. This will be $M_1(0) = 1$ and $M_1(0) = 0$.

(b) Simulate the regression coefficients of M_2 by making a draw from a **MVN** distribution parametrised by regression coefficients of M_2 and the variance-covariance matrix of those coefficients.

(c) Use the mediation formula to determine which predicted probabilities to compute based on the combinations of M_1. This will be:

$$M_2(0, M_1(0) = 1)$$

$$M_2(0, M_1(0) = 0)$$

(d) Simulate the necessary M_2 and use these in the mediation formula:

$$\begin{aligned} M_2(0, M_1(0)) = \Pr(M_1(0) = 1) \cdot M_2(0, M_1(0) = 1) + \\ \Pr(M_1(0) = 0) \cdot M_2(0, M_1(0) = 0) \end{aligned}$$

(e) At the end of each simulation, we will have a vector of simulated predicted probabilities of M_2, one for each subject.

6. Estimate $Y(1, M_1(1), M_2(0, M_1(0)))$.

(a) Identify the combinations of $M_1(1)$ and $M_2(0, M_1(0))$. This will be:

$$M_1(1) = 0, M_2(0, M_1(0)) = 0$$

$$M_1(1) = 0, M_2(0, M_1(0)) = 1$$

$$M_1(1) = 1, M_2(0, M_1(0)) = 0$$

$$M_1(1) = 1, M_2(0, M_1(0)) = 1$$

(b) Simulate the regression coefficients of Y by making a draw from a **MVN** distribution parametrised by regression coefficients of Y and the variance-covariance matrix of those coefficients.

(c) Use the mediation formula to determine which predicted probabilities to compute

based on the combinations of M_1 and M_2. This will be:

$$Y(1, M_1(1) = 0, M_2(0, M_1(0)) = 0)$$
$$Y(1, M_1(1) = 0, M_2(0, M_1(0)) = 1)$$
$$Y(1, M_1(1) = 1, M_2(0, M_1(0)) = 0)$$
$$Y(1, M_1(1) = 1, M_2(0, M_1(0)) = 1)$$

(d) Simulate $Y(1, M_1(1), M_2(0, M_1(0)))$ by applying the mediation formula:

$$Y(1, M_1(1), M_2(0, M_1(0))) =$$
$$Y(1, M_1(1) = 0, M_2(0, M_1(0)) = 0) \cdot \Pr(M_1(1) = 0) \cdot \Pr(M_2(0, M_1(0)) = 0) +$$
$$Y(1, M_1(1) = 0, M_2(0, M_1(0)) = 1) \cdot \Pr(M_1(1) = 0) \cdot \Pr(M_2(0, M_1(0)) = 1) +$$
$$Y(1, M_1(1) = 1, M_2(0, M_1(0)) = 0) \cdot \Pr(M_1(1) = 1) \cdot \Pr(M_2(0, M_1(0)) = 0) +$$
$$Y(1, M_1(1) = 1, M_2(0, M_1(0)) = 1) \cdot \Pr(M_1(1) = 1) \cdot \Pr(M_2(0, M_1(0)) = 1)$$

(e) At the end of each simulation, we will have a vector of predicted probabilities of Y **PO** and in this case $Y(1, M_1(1), M_2(0, M_1(0)))$.

(f) We then take the expectation of the simulated expected probabilities for each simulation.

The procedure outlined is used to simulate a single Y **PO**. We will then need to repeat this for the **PO**s required to compute the causal effect. Once we have simulated both **PO**s, we can then obtain an estimate of the **OR** by using the expectation of the Y **PO**s across all simulations. This thus concludes the extension of the simulation method for two mediators.

4.5.5 Variation V: Simulate regression coefficients with bootstrap.

The last variation is the use of bootstrap for the simulation for the regression coefficients. In lieu of making a draw from a **MVN** distribution, one can instead use a bootstrap to simulate the regression coefficients. This applies regardless of which **PO** one is simulating. Using the **PO** $Y(1)$ as an example, the process of simulating the regression coefficients using bootstrap are as follows:

1. Sample n subjects from the model matrix X_Y with replacement to form a new model matrix.

2. Use the new, re-sampled, model matrix to estimate the model Y.
3. Use the regression coefficients of this newly estimated model as a single simulate of the regression coefficient of the model Y.

The bootstrap is ideal if one does not wish to make the assumption of **MVN** of the estimated regression coefficients. This is applicable particularly when the sample size is small. The drawbacks of the bootstrap is that it is much more computationally intensive than the **MVN** approach and the **MVN** approach is also statistically more powerful i.e. for a given sample size, using the **MVN** approach allows one to detect smaller causal effects compared to the bootstrap approach. It is important to note that the aim of the simulation is to reproduce the sampling distribution of the outcome and both the bootstrap and the **MVN** method are valid ways to do this.

Another drawback of the bootstrap approach come when one is estimating logistic regressions. If we were to refit a logistic model at each run of the simulation, we can run into convergence problems during the refits particularly when the probability of the outcome is low to begin with. With a low probability, there is increased likelihood that the proportion of re-sampled outcome being 1 may be too low to make estimation of the parameters possible. One might choose to discard the model matrix and sample a new one. Doing so however implies that only certain combinations of the resampling will be used and thus the probability of selecting each combination of subjects is not uniform. This violates a basic assumption of the non-parametric bootstrap and while there are discussions of such a phenomenon (E.g. Gomes and Oliveira (2001)) there is currently no consensus on how such a situation is to be resolved.

4.6 Precision of estimates from simulation-based estimator

The simulation-based estimator differs in one important respect from the model-based estimator. The *precision* of the simulation-based estimator is highly dependent on the number of simulations used. *Precision* in this context refers to the discrepancies between the model- and simulation-based estimators. The model-based estimate in this context is being used as a gold standard, a deviation from traditional methods of assessing estimators. This deviation is borne from the idea that both estimators were constructed using the same theoretical background and since they

both draw upon the same theoretical background and are built using the same set of models for outcomes and mediators, at best, the simulation-based estimator is expected to reproduce the estimates from the model-based estimator should it be possible to use the model-based estimator to conduct the estimation. Further discussion on comparisons between the model- and simulation based estimators can be found in the next chapter.

The *precision* of the estimates is closely linked to the number of simulations used to obtain the estimates. Generally, the larger the number of simulations, the more reliable the estimates will be at the cost of lengthier computation times. This increase can stretch to hours or even days for larger, more complex models. This increase in computational times greatly increases if the results for the sensitivity analyses are required because the same set of simulations will be conducted for each level of confounding tested in the sensitivity analysis. At some point however, further increases in the number of simulations will only yield negligible improvements in precision, a case of diminishing returns. The number of simulations at this point strikes an optimum balance between precision and computation times. Unfortunately, the only way to identify this point is through an iterative trial and error procedure which begins by first obtaining an estimate using a predetermined number of simulations x. With this initial estimate, we then iterate over the following steps:

1. Increase the number of simulations by y amount.
2. Obtain an estimate using $x + y$ simulations.
3. If the estimates obtained with x simulations and $x + y$ simulations differ by z amount, go to step 1 and repeat the steps.

The values of x, y and z need to be predetermined. Lastly, this procedure needs a breaking condition to guard against excessively long computation times which can be triggered by having a very small z. This procedure outlines the optimal way to set the number of simulations but is not currently implemented. Instead, a default value of 100000 is used and in practice, this number of simulations had been found to be sufficiently precise. This value can be overridden should there be a need to do so.

4.7 Summary

In summary, this chapter presented a solution to three methodological gaps for the estimation of mediated causal effects. Firstly, there was a lack of expressions for estimators in multiple sequential mediation that allowed the testing of path specific hypotheses. Secondly, there was also a lack of expressions for causal **OR** estimators for both the single and multiple mediator scenarios which does not rely on restrictive assumptions. Thirdly, a solution is needed for evaluating these expressions in order to obtain causal effect estimates. The lack of expressions for the estimators in the first two methodological gaps was bridged by building upon and extending the *mediation formula*. The efforts yielded solutions for expressions for estimators for single and multiple sequential mediation for both causal differences and causal **OR**. Besides providing solutions to scenarios where there were none previously, these solutions provided also differed from previous efforts in that they allowed the testing of path specific hypotheses.

These expressions then need to be evaluated in order to obtain an estimate of the causal effect. While many of the expressions after simplifying have closed form solutions whereby one can substitute in estimated model parameters to obtain an estimate of the causal effect, many do not. This is primarily due to an integral that has no closed form solution. The second part of this chapter discussed how a quasi-Bayesian simulation method had been used to evaluate these expressions. This allowed one to estimate mediated causal effects for any combination of mediators and outcomes for both causal differences and causal **OR**. Using the simulation method to estimate the causal **OR** represented a significant challenge due primarily to the non-collapsability of **OR** and special attention had to be paid to how the simulations were summarised. The relationship between the number of simulations and the precision of the estimates was also discussed with a proposed algorithm to determine the optimal number of simulations. This algorithm however is not currently implemented and a default value of 10000 is adopted as the default. This number of simulations had been found to be sufficiently precise in practice.

Lastly, these were implemented in an *R* programme, *seq-med* chapter G and in the next chapter, several checks will be presented to ensure that the estimates are consistent with causal estimates obtained using other methods for the scenarios where there were prior methods that could be

used. An application of this package to an **RCT** will also be presented.

Chapter 5

Validation and application of the novel causal effects estimator

5.1 Introduction

In the last chapter, I discussed the theoretical foundation and implementation details of a novel causal effects estimator. This novel causal effects estimator enables the estimation of causal effects under two scenarios. Firstly, it allows the estimation of causal odds ratio when a continuous mediator is nested within a binary outcome or another mediator. Secondly, it enables the estimation of sequentially mediated causal effects in the presence of more than a single mediator. A mix of the two scenarios is also addressed with this novel estimator, i.e. sequentially mediated causal effects with continuous potential outcome (**PO**) nested within a binary outcome or another mediator. While the theoretical foundation provides strong justification for the existence of the causal effect estimator, it does not cover an implementation of the causal effect estimator. Therefore, in order to validate that the implemented estimator performs within expectation, a validation study needs to be conducted. This validation study seeks to answer two questions. Firstly, are the estimated causal effects in line with expected causal effect estimates according to the theoretical foundation? Secondly, if there are discrepancies, should these be of concern and are these discrepancies likely to change the estimates in a meaningful manner?

Focusing on the first question that the validation check seeks to answer, what is 'in line with

expected causal effect estimates' needs defining. The theoretical foundation upon which the novel estimator depends on is derived from an extension of the mediation formula developed within this thesis. The model-based estimator similarly draws upon the mediation formula for its theoretical foundation. The extension of the mediation formula additionally allowed the model-based estimator to estimate sequentially mediated causal effects. This shared theoretical foundation implies that given the same inputs, they should produce the same causal estimates. Any discrepancies are highly likely to be attributed to implementation differences and this is further explored in this chapter. Implementation differences here refer to how the estimation method was implemented in software. The validation check uses the similarity between the model- and simulation-based estimators to determine whether or not the novel simulation-based estimator produces the same estimates as the model-based estimator for the same set of inputs. For the scenarios where both the model- and simulation-based estimator could be used to estimate the mediated causal effects, the estimates from both estimators would be compared with each other. This forms the primary way that the simulation-based estimator is validated for the scenarios where the model-based estimator could also be used to conduct the estimation.

While the model- and simulation-based estimators share similarities and both could be used to estimate mediated causal effects in a number of scenarios, there are scenarios where only the simulation-based estimator could be used. This was one of the reasons the simulation-based estimator was developed in the first place. For the scenarios where only the simulation-based estimator could be used to estimate the mediated causal effects, I relied on a unique property of the total effects (**TE**) to conduct the validation. Recall in eqs. (3.4.5) and (3.4.6) that the **TE** can be estimated either using the model for the outcome without any mediators, or it can be estimated using the outcome model that contains mediators along with the models for the mediators. The two methods of estimation the **TE** would subsequently be referred to as the method using the outcome model only and the method using the outcome with mediators model. Theoretically, both methods of estimating the **TE** should yield the same causal estimate. However, in practice, there is a larger number of mathematical operations used in the latter method of estimation and depending on the type of mathematical operations incurred, there could be appreciable precision loss. This precision loss, if present, would then make the two methods of estimation produce estimates that differ by the precision loss. Generally, logarithmic and exponentiating operations incur a larger loss of precision compared to sums and products and this precision loss can be quantified by making a comparison of the **TE** estimated using the above two methods of

estimating the **TE**. This will be explored in greater detail in the next section.

As discussed previously, there are two ways to estimate the **TE**. The method of estimation using the outcome model only can utilise both the model- and simulation-based estimator. For the method of estimation using the outcome with mediators model, the simulation-based estimator can be used for all scenarios but this is not so for the model-based estimator. This is because this second method of estimating the **TE** uses the same mechanism as that used for estimating the mediated causal effect. Therefore if under certain scenarios mediated causal effects cannot be estimated using the model-based estimator, likewise under these scenarios the method of estimating **TE** using the outcome model with mediators method cannot be used. The ability to estimate **TE** with the model-based estimator for all scenarios using the outcome model only however offers us a way to validate the simulation-based estimator for the scenarios where the model-based estimator cannot be used to estimate the mediated causal effects. For these scenarios, we can use the model-based estimator to estimate the **TE**s and compare these estimates to the estimates of **TE** from the simulation-based estimator using the outcome model with mediators to conduct the estimation. The method of estimating the **TE** using the outcome model with mediators for the simulation-based estimator follows the same procedure as estimating mediated causal effects. The only difference between estimating the **TE** and mediated causal effects is different **PO** definitions, i.e. different values of r in the definition of the causal effect. Given this similarity, a validation of the **TE** estimate using the outcome model with mediators method would thus be a validation of the simulation-based method of estimating mediated causal effects.

Another important thing to note about the validation study is its deviation from traditional ways in which new estimators are assessed. Typically, new estimators are assessed for their bias and variance with respect to a true value. This true value is used to generate data for a population. From this population, a sample is drawn and the novel estimator is applied to determine the bias and variance of the estimator with respect to the true value and the sample size. In such a study, the manner in which the data is generated is of prime importance because if the data generated deviates from the predetermined true value, then any assessment of bias and variance can and will be erroneous.

This validation study deviates from this established practice for two reasons. Firstly, the

simulation-based estimator shares a theoretical base with an established estimator, the model-based estimator. This similarity between the two estimators imply that they should produce very similar estimates. Any deviations that cannot be appropriately explained implies an error in the implementation of the simulation-based estimator. Secondly, if the established practice had been used, the results would quantify the bias and variance of the novel estimator. This bias and variance would be the same between the model- and simulation-based estimator since they both share the same theory base provided that the simulation-based estimator had been implemented correctly. The quantification of the bias and variance alone without any reference to the model-based estimator does not provide adequate information about whether the simulation-based estimator has been implemented correctly. If the bias and variance of both estimators were quantified and compared, there is still a lack of direct comparison between the two estimators which would provide the strongest indication of whether or not implementation errors exist in the simulation-based estimator. It is for these two reasons that this validation study deviates from established practice in order to more robustly identify any implementation errors present in the simulation-based estimator. Also arising from this deviation is that the manner in which the data is generated is of less importance because regardless of how the data is generated, both estimation methods use the same set of data with the resulting estimates being compared. This same comparison guided the development of the simulation-based estimator along the way, and revealed mistakes in the programming of the estimator which were rectified when the mistakes were found.

Once the implementation has been validated, the second part of this chapter will demonstrate the use of the novel causal effect estimator on a real trial, the Carers' Assessment, Skills and Information Sharing (CASIS) trial (Goddard et al., 2013). The demonstration highlights the use of the novel estimator and the information about the estimates that it provides.

5.2 Validation study: Methods

In the introduction, a general outline of the validation of the novel estimator was presented. In this section, I will go into the detailed setup of the study. The validation study is split into two parts. The first part investigates the impact of precision loss associated with an increase in the number of mathematical operations used to estimate the causal effects. This quantification

of the precision loss allows us to consider if discrepancies between model- and simulation based estimates fall within the range of precision loss. If it falls within the range of precision loss and there are known increase in number of mathematical operations, the discrepancy is likely attributable to the increased number of mathematical operations. This quantification of precision loss thus forms the context within which to consider the second part of the study which compares the estimates obtained from the model- and simulation-based estimators.

For the first part of the study, the precision loss is investigated by using the model-based estimator in scenarios where it can be used to estimate the causal effects. This allows one to estimate the **TE** using the model-based estimator in two ways. The first way uses the outcome model without mediators and the second way uses the outcome model with mediators included. These are discussed in chapter 3 and the relevant equations are eqs. (3.4.5) and (3.4.6). The **TE** estimated from both approaches should yield the same estimates theoretically. In practice however, due to the difference in the number of mathematical operations, the causal effect estimates differ. By using both approaches to estimate the total effects, we can have some quantification of the degree to which the precision loss occurs.

The second part of the study is split into two further parts. The first compares the model- and the simulation-based estimates for which there are solutions for estimating the mediated causal effects using the model-based estimator. The second part compares the **TE** between the two estimators for scenarios where a model-based estimate for the mediated causal effect does not exist. I have named the scenarios to reflect the combination of type of outcomes, second mediator and first mediator in that order. The first mediator is nested within the second mediator and both the first and second mediators are nested within the outcome. *B* indicates a binary variable modelled using logistic regression while *C* indicates a continuous variable modelled using linear regression. Therefore *CBC* represents the combination of a continuous outcome, a binary second mediator, and a continuous first mediator. The combination *BBC* represents a binary outcome, a binary second mediator and a continuous first mediator. The scenarios for which a model-based estimate for mediated causal effects exists are *BBB*, *CBB*, *CCB*, and *CCC*. The scenarios where the model-based estimator cannot be used to estimate mediated causal effects are *BBC*, *BCB*, *BCC*, and *CBC*.

For the first set of model combinations, the mediated causal effects from both the model- and

simulation-based estimators are compared with each other and for the second set, only the **TE**s are compared due to a lack of mediated causal effects from the model-based estimators. The **TE** estimated by the model-based estimators uses the outcome model without any of the mediators and theoretically, this is a like for like comparison since both the model- and simulation-based estimators are both estimating the same defined quantity. In practice however, due to a large increase in number of mathematical operations that the simulation-based estimator incurs, it is expected to differ from the model-based estimator owing to precision loss. This precision loss was investigated as part of the validation study and the results from this second part of the validation study should be considered in the context of the expected precision loss from the first part.

For each set of model combinations, the causal effects are estimated across a variety of scenarios aimed at representing commonly encountered situations found randomised controlled trials (**RCT**s) of mental health interventions. These scenarios varied sample and effect sizes. The sample sizes used were 124, 250, 500, and 1000. Each subsequent sample size is double of the previous one except for 250. The previous sample size used is 124 instead of 125 to ensure equal numbers of subjects offered each of the treatment and control arm. The effect sizes are defined using the Cohen's D (Cohen, 1988) and the sizes used were small, medium, and large corresponding to effect sizes of 0.8, 0.5 and 0.2. This effect size relates only to the **TE**s. For continuous outcomes, the control treatment is set to have a mean effect of 0. For binary outcomes, the Cohen's D is undefined for causal odds ratio (**OR**) so in order to maintain the same validation parameters across models, an adaptation to the Cohen's D was used. This adaptation treats the predicted probability of the binary outcome as a continuous variable and by setting the control treatment to have a mean predicted probability of the outcome of 0.5, the predicted probability for the experimental treatment corresponding to a Cohen's D for small, medium, and large effect sizes can then be computed. The computed mean predicted probabilities are 0.599 for small, 0.736 for medium, and 0.848 for a large effect. These correspond to **OR**s of 1.49, 2.79, and 5.58 respectively.

So in total, for eight different model combinations, across four sample sizes and three effect sizes, I explored a total of 96 combinations. For each combination, 100 datasets were generated for a total of 9600 datasets each consisting of an outcome, two mediators, a treatment variable and a covariate. Each dataset was seeded with a different seed to ensure that no two datasets

were alike. The respective models, linear regressions for continuous dependent variables and logistic regression for binary dependent variables, were fitted for the outcome and mediators for each dataset. The model- and simulation-based estimators were then applied to each of models. For 9600 sets of models with two estimators, a total of 19200 estimation procedures were conducted. A modest 10000 bootstraps were used for the model-based estimator and the same number of simulations were used for the simulation-based estimator. This is a magnitude short of the default value of 100000 but is used to keep the validation study manageable in terms of time to run the simulations.. This default value was derived through a process of trial and error and represents the number of simulations that strikes a balance between computation times and the precision of the estimates. If the default value of 100000 simulations was used, the time required to complete the simulations would be about one and a half months compared to about a week by reducing the number of simulations to 10000. More importantly, the increased number of bootstraps and simulations are unlikely to provide us with additional insight into the behaviours of the estimators since the additional simulations are expected to improve accuracy at the third decimal place and beyond. An improvement in precision at such magnitudes does not change the conclusions of the validation study. Lastly, the mean of the results for each of the 96 combinations across 100 datasets are taken and presented as graphs in the following section.

5.3 Validation study: Results

The results of the validation checks are presented in two parts. The first being the investigation of the precision loss and the second being the comparison of the estimates obtained from the model- and simulation-based estimators.

5.3.1 Precision loss

This part of the validation study forms the context from which to interpret the results of the comparisons between the model- and simulation-based estimates. It informs us of the precision loss due to an increased number of mathematical operations and does so primarily by comparing the **TE** estimates from the model-based estimator computed in two different ways.

The results are presented in fig. 5.1 for the continuous outcomes and fig. 5.2 for the binary outcomes. In both sets of graphs, the estimate of **TE** estimated using the outcome with the mediators together with the mediator models are on the left represented by a circle and the **TE** obtained using only the outcome model without the mediators is on the right represented by a triangle. The bars of each estimate represent the confidence intervals. A dotted line on each of the graphs represents the null causal effect value. This dotted line is at $Y = 0$ for continuous outcomes where the causal effects estimated are differences and is at $Y = 1$ for binary outcomes where the causal effects estimated are **OR**. The results are split into continuous and binary outcomes because the computation of each uses different mathematical operations, thereby allowing us to determine if the type of mathematical operations play a role in the precision loss. For the continuous outcomes, the main mathematical operations involved are sums and products of numbers. Focusing on the differences between the two ways that the **TE** were estimated, we can see that the estimates and confidence intervals line up almost exactly fig. 5.1. Any differences are negligible. Next, we move on to the binary outcomes. In the estimation of causal **OR**, mathematical operations involved include sums, products, and taking of natural logs and exponents. Only a single combination with binary outcomes, *BBB*, has a model-based solution using the method involving the outcome model with mediators. Focusing on the differences in estimates, it is largest for the largest effect size with the smallest sample size. This difference however is not concerning since the confidence interval is much larger. This difference reduces with a smaller effect size and larger sample size. This suggests that the discrepancy scales with the size of the effect where a larger effect size has a larger difference. A larger effect size indicates a higher mean predicted probability of the outcome being 1 for the experimental group with the mean predicted probability for the control group being held at 0.5. In order to obtain the mean predicted probabilities, the inverse link function of the logistic regression needs to be applied to the sum-product of the estimated regression coefficients with the dataset. In the process of conducting the estimation, logarithmic and exponent operations incurred in the link function and inverse link function of the logistic regression is simpler for the estimation of **TE** using the outcome model without mediators since the sum-product of fewer terms needed to be computed before applying the inverse link function of the logistic regression. This reduction in computation means that any precision loss is reduced owing to a reduction of the terms represented in the computer. This relates to how a computer represents numbers with decimals internally and it does so not by storing it precisely but as an approximation with a high degree of

Figure 5.1: Comparison of total effect causal estimates using models with and without mediators for continuous outcomes

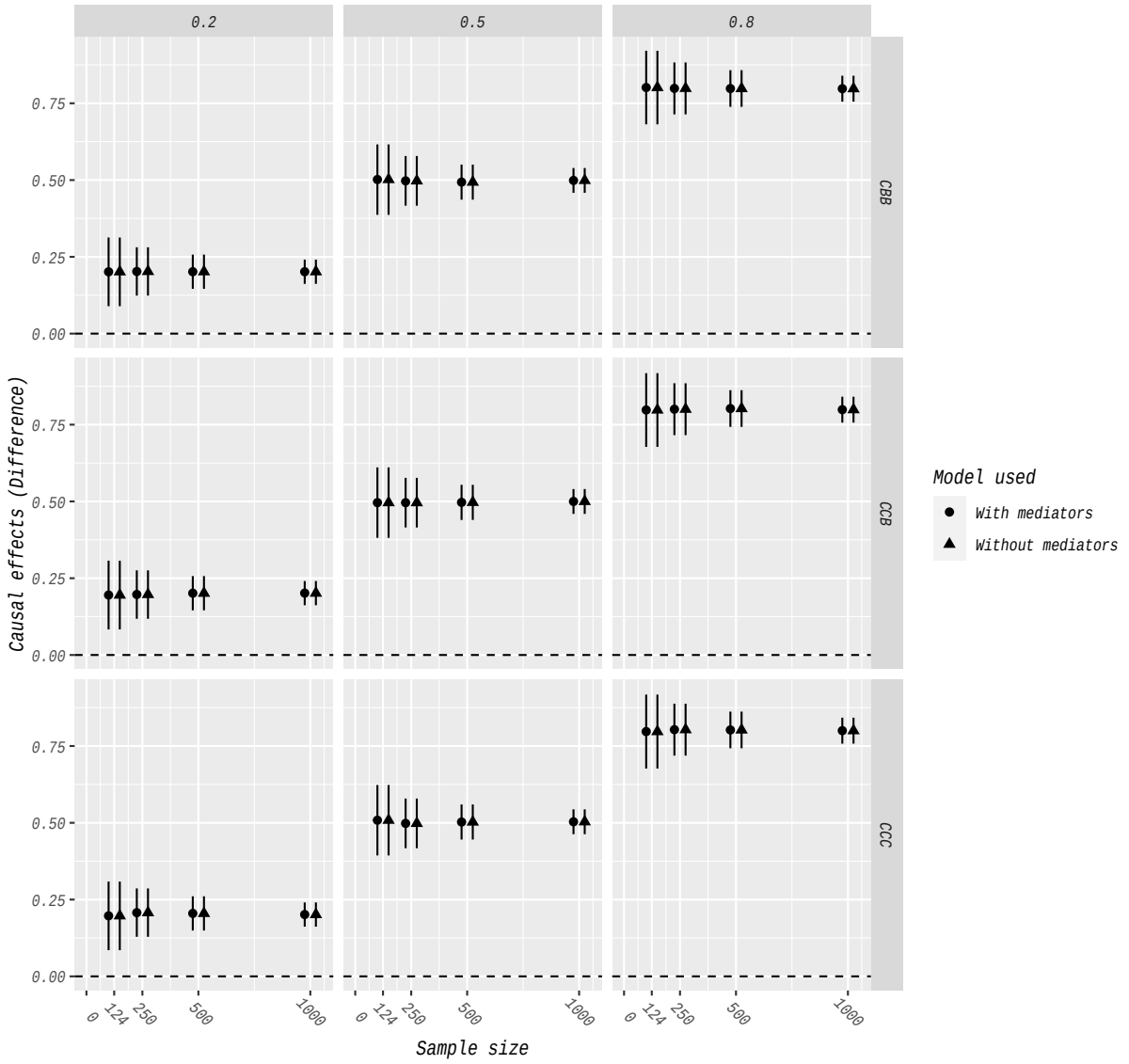
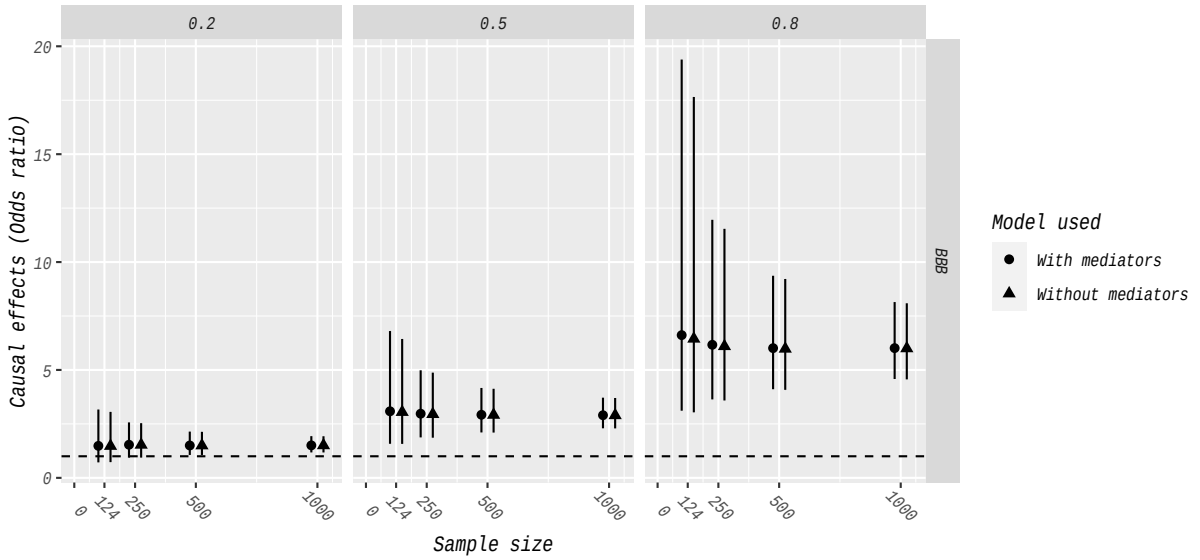


Figure 5.2: Comparison of total effect causal estimates using models with and without mediators for binary outcomes



accuracy. This accuracy however degrades with increased number of computations. This meant that for a larger effect size, operations incurred in the computation, particularly computations such as logs and exponents, incur a larger precision loss. This accounts for the observation of differences in larger effect sizes. This difference however reduces with a larger sample size even if the effect size remains large likely due to cancellation of precision losses. A larger sample size makes it more likely for a precision loss to be offset by another precision loss in the opposite direction.

Furthermore, the causal effect used for binary outcomes, the causal **OR**, uses the predicted mean probability of each of the control and experimental group twice in the computation of the causal **OR** (each odds within the ratio uses the predicted probability twice). This repeated use of an estimate with some lack of precision likely also made precision issues more perceptible for the causal **OR** as compared to a continuous outcome. Lastly, it should be noted that a small change in the mean predicted probability will result in a seemingly large change in **OR**. The largest precision loss in this part of the study for causal effects defined as a difference is on the third decimal place whereas for the causal **OR** is on the second decimal place. This is an order of magnitude in difference and is most likely due to the way the causal **OR** is constructed, as a ratio between two odds. As an example, consider a predicted mean probability of 0.5 and 0.8 for the control and experimental groups respectively. The **OR** computed from both the probabilities

is 4. If the mean predicted probability for the experimental group was 0.81 instead, the **OR** is 4.26. A difference of 0.01 probability in the experimental group results in a difference of 0.26 in **OR**. Whether or not this is an important difference has to be decided by the researcher, but it is important to note that in a study with large effects and small sample sizes, a small difference in predicted probability is more likely to lead to precision loss of concern. As shown in this sub-study, this precision loss is not a unique problem for simulation based estimators since this sub-study uses model-based estimators only. The magnitude of this precision loss provides the context within which to consider the discrepancies between model-based and simulation-based estimators with the latter expected to have a higher level of precision loss owing to the much larger number of operations incurred.

5.3.2 Comparison of model- and simulation-based estimators

This second part of the validation study is split into two parts. The first part addresses combinations of models with both model- and simulation-based solutions available and the second part addresses the models where only simulation-based estimators are available for mediated causal effects. The first part of the study compares the **TE**, and direct and mediated causal effects between the model- and simulation-based estimators. The second part, because of the lack of estimated mediated causal effects for model-based estimators, only compares the **TE** between the two estimators as mentioned in section 5.2.

The results for the first part are presented in figs. 5.3 to 5.6 for the model combinations *BBB*, *CCB*, *CBB*, and *CCC* respectively and each figure is a 3×5 array of graphs. The rows of the array represent different effect sizes from the smallest on the left and largest on the right. The columns of the array represents the causal effects, total, direct, indirect through $M1$, indirect through $M2$, and indirect through both $M1$ and $M2$ from top to bottom. Within each cell of the array is a graph comparing the causal effect estimates for different sample sizes. The model-based estimates are represented as circles and the simulation-based estimates as triangles. The y-axis for the **TE**s and direct effects (**DE**s) share the same scale while the indirect effects share a different scale. The reason for this is because the effects are different in magnitudes and if they were to all share the same scale, the confidence interval of the indirect effects (**IE**s) will not be perceptible. Of the four model combinations, only a single model combination, *BBB*,

Figure 5.3: Validation checks for *BBB* models

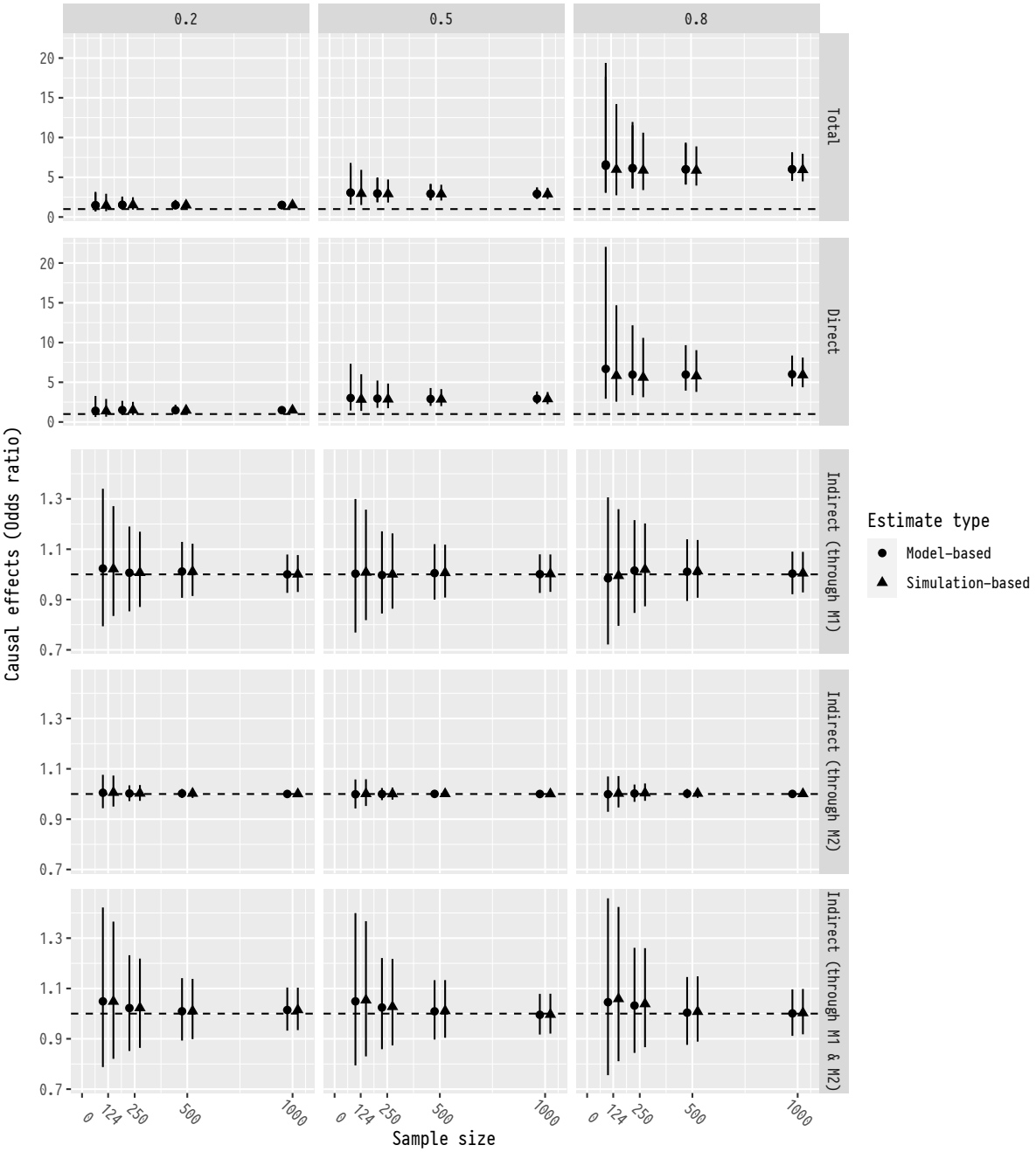


Figure 5.4: Validation checks for *CBB* models

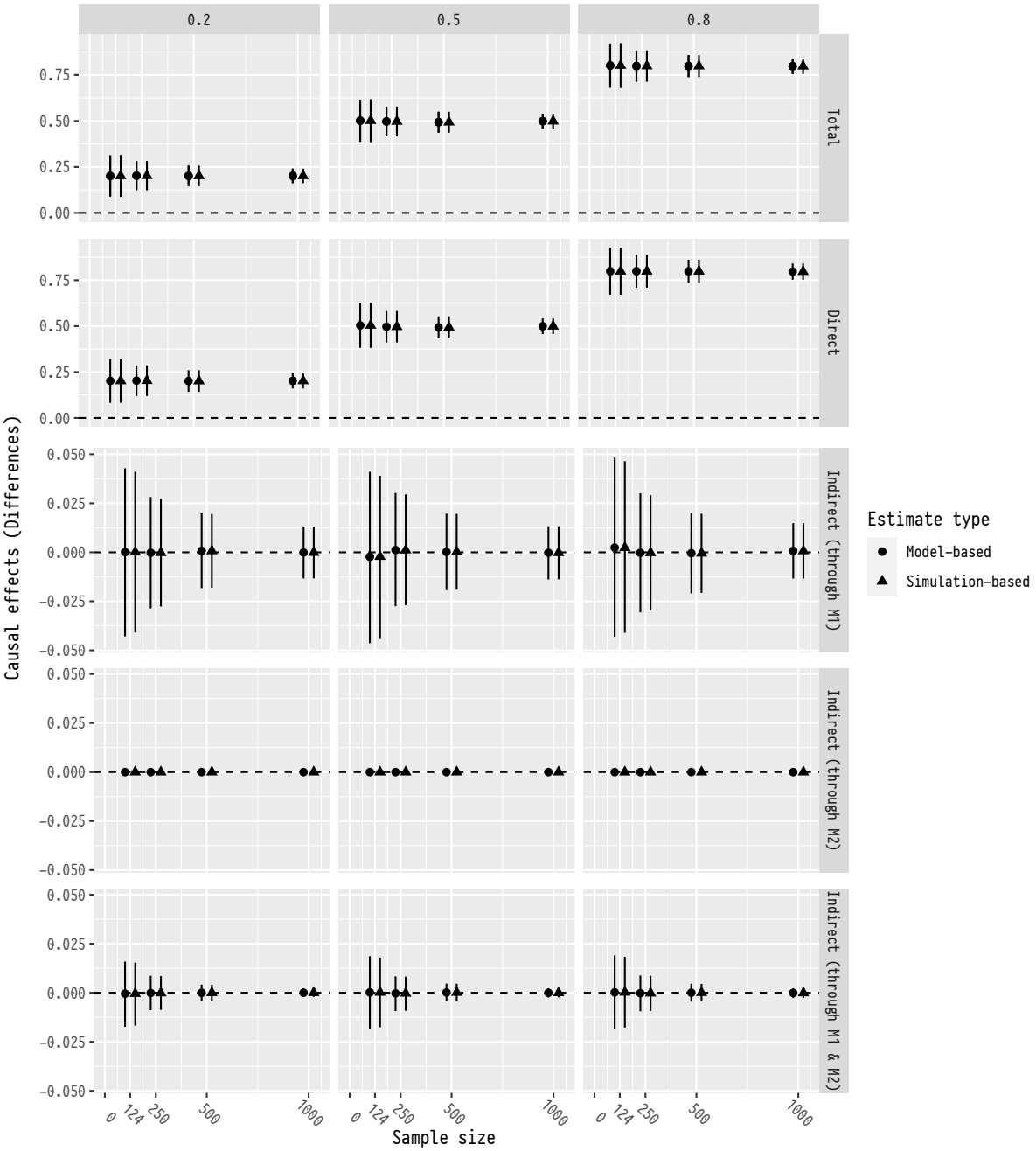


Figure 5.5: Validation checks for *CCB* models

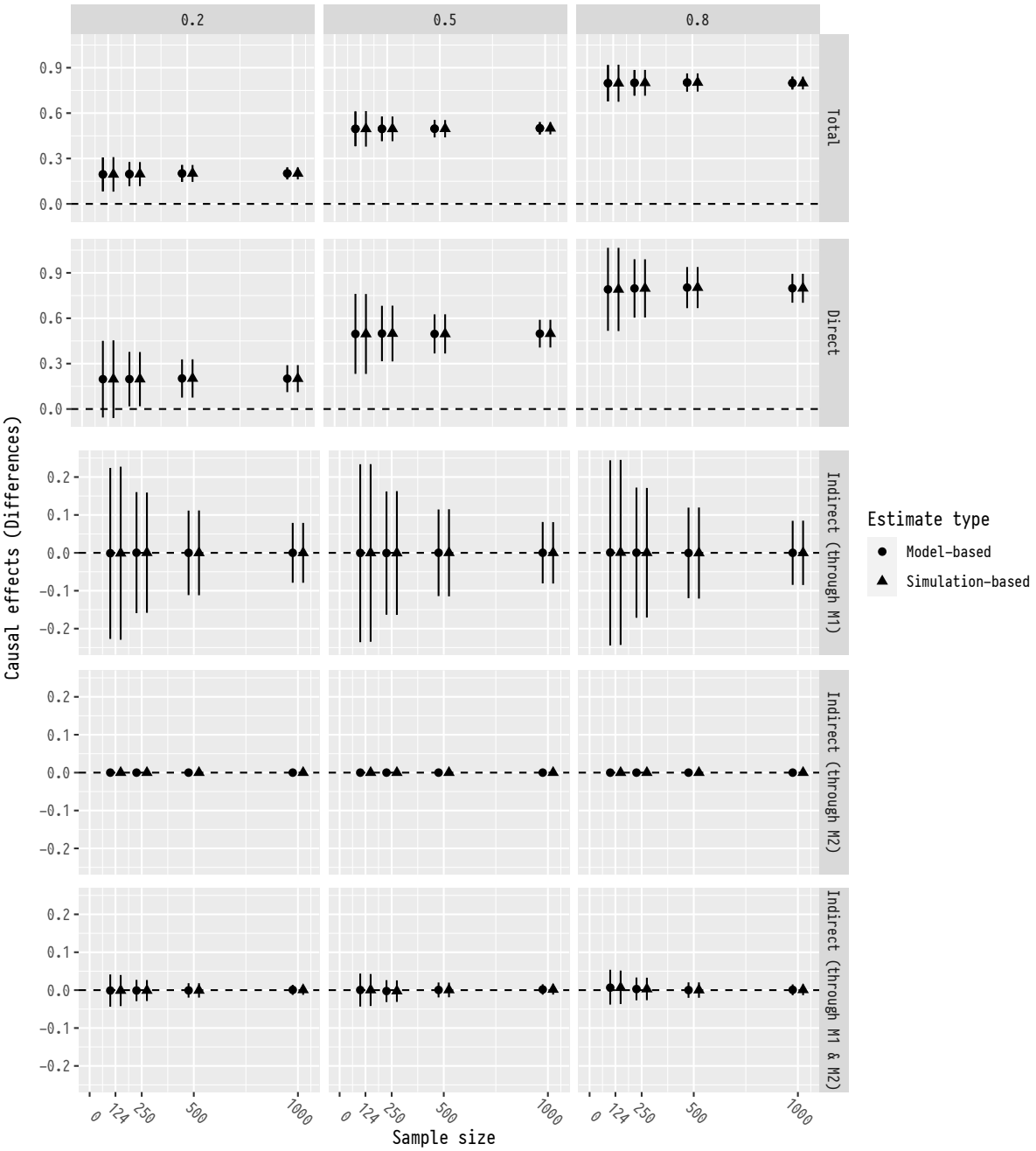
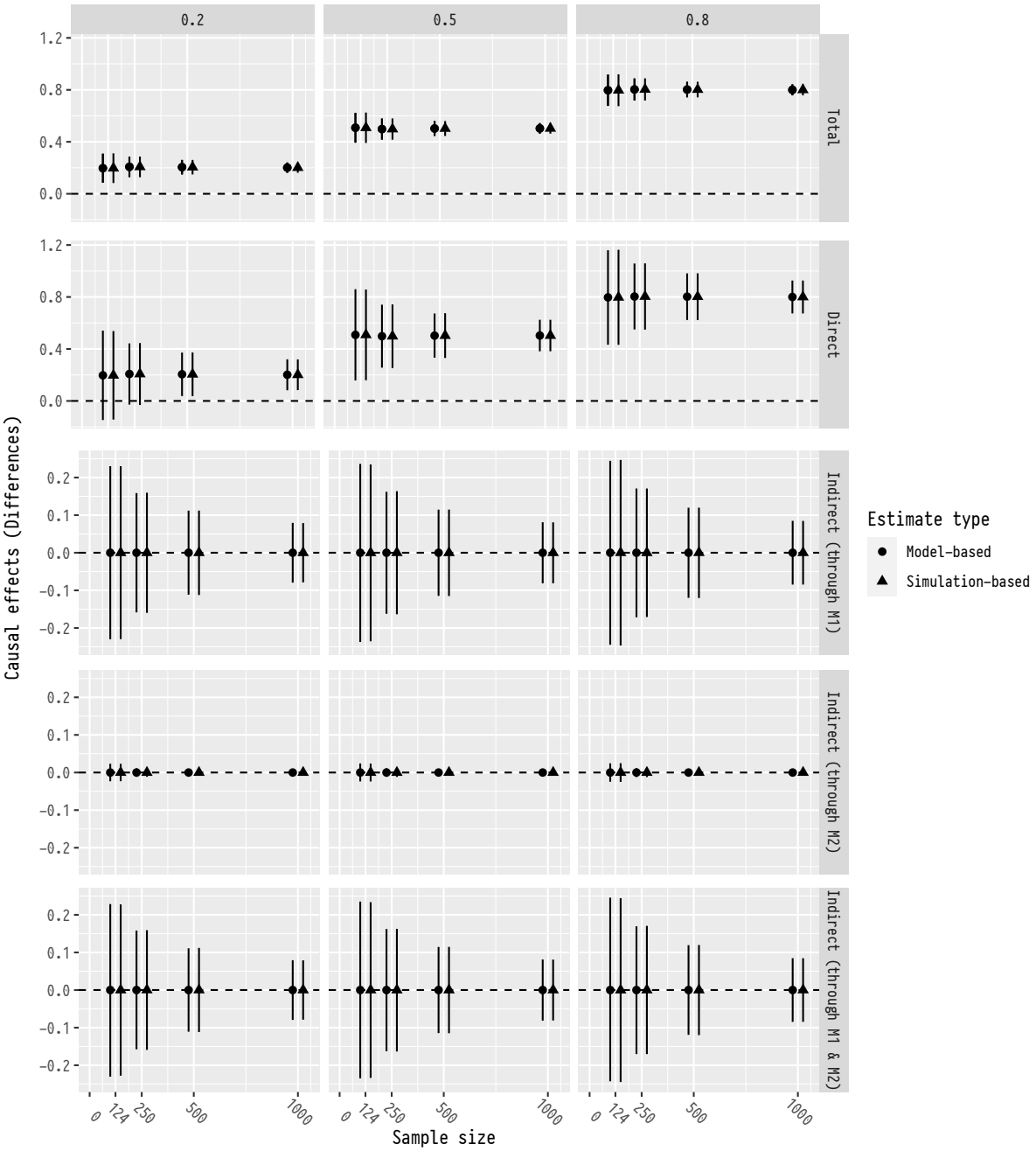


Figure 5.6: Validation checks for *CCC* models



has a binary outcome. The rest of the model combinations use continuous outcomes. For the continuous outcomes, all the causal effect estimates and confidence intervals for the model- and simulation based estimators are almost identical in magnitude. The confidence intervals of all the estimates (both continuous and binary outcomes) line up with expectations with the largest confidence intervals for the smallest sample size and the smallest confidence intervals for the largest sample sizes. The indirect effects also exhibit asymmetrical confidence intervals which is to be expected given the way the confidence intervals are generated by ranking all the simulated estimates and then reading off the required percentiles, similar to what Imai et al. (2010) did in the *mediation* package.

For the binary outcome combination, *BBB*, the estimates demonstrate the same patterns as that observed in the sub-study on precision loss, i.e. larger differences for smaller sample sizes and larger effect sizes. This suggests that the differences observed between the model- and simulation-based estimate fall within the differences expected when there is precision loss. This similarly occurs for causal **OR** and this difference is largest for large effect sizes and smallest for large sample sizes. Given that the simulation-based estimators incur more mathematical operations, and in particular operations involving taking the log and exponentiating numbers, the most likely explanation is that this difference is due to precision loss as was investigated previously. When considered within the width of the confidence intervals, the difference is small and unlikely to alter the interpretation of the results.

Lastly, it should be noted that for the smaller sample sizes, the confidence intervals for the model-based estimates are larger than the simulation-based estimates. This is due to the use of 10000 rather than the default value of 100000 bootstraps used for the model-based estimators. This difference is minimal once the default value is used.

Next, I will discuss the results for when there are no model-based estimates for mediated causal effects. These results for the model combinations *BBC*, *BCB*, *BCC*, and *CBC* are shown in figs. 5.7 to 5.10. Here, similar patterns are observed with the causal **OR** estimates exhibiting a bigger difference than the causal difference estimates between the model- and simulator-based estimates. The causal **OR** differences is again larger for larger effect sizes and small sample sizes. This is likely due to precision loss as explained previously. Overall, the validation checks provide strong indications that the simulation-based estimator was implemented correctly. Any

Figure 5.7: Validation checks for *BBC* models

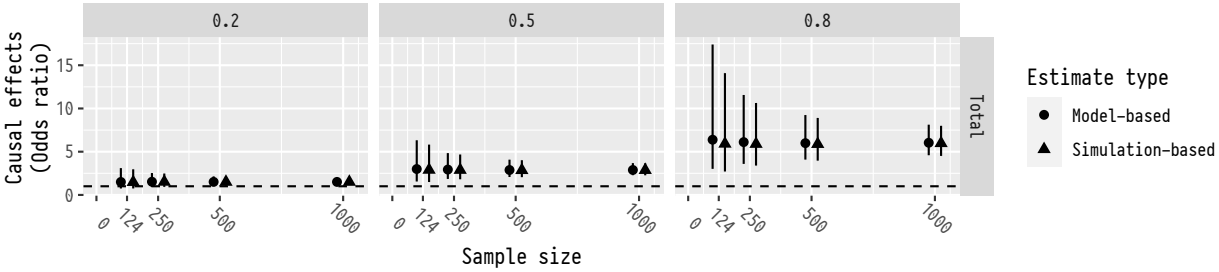


Figure 5.8: Validation checks for *BCB* models

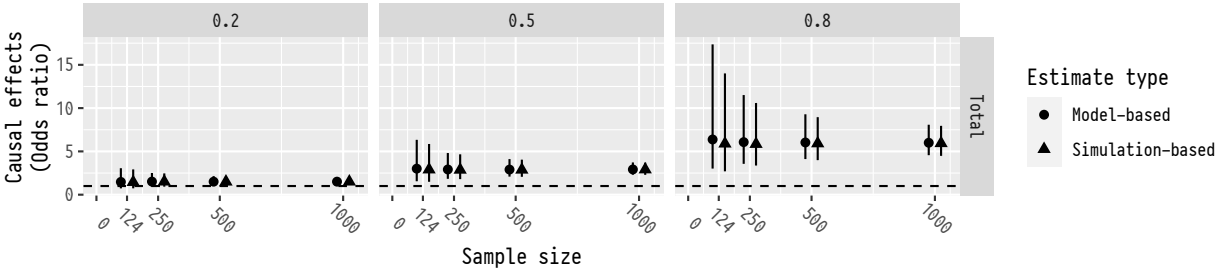


Figure 5.9: Validation checks for *BCC* models

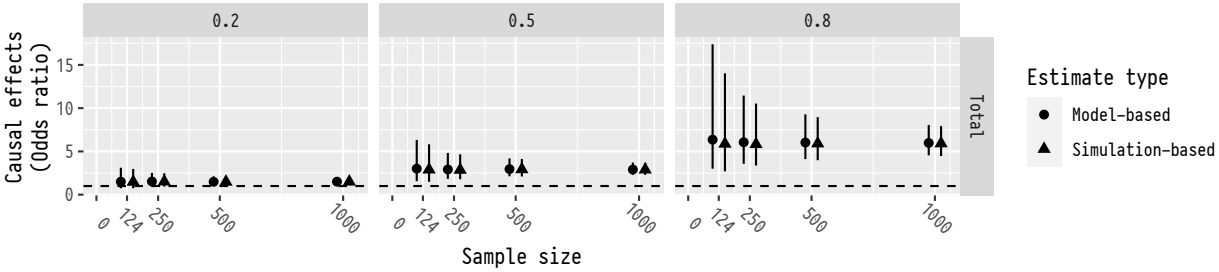
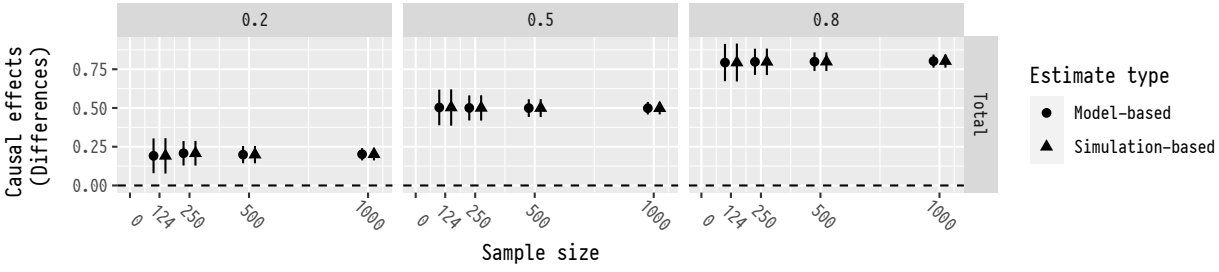


Figure 5.10: Validation checks for *CBC* models



observed differences were likely due to precision loss and when considered with the width of their respective confidence intervals, appear unlikely to alter the interpretation of the findings of a study. The next section focuses on an application of this novel estimator to a real **RCT** to demonstrate its real world utility.

5.4 Application of estimator to CASIS

The **CASIS** trial was introduced previously (section 3.3). As a recap, the **CASIS** trial was an **RCT** that primarily aimed to assess the efficacy of a set of education curriculum for the caregivers of patients with eating disorders (**ED**) (Goddard et al., 2013). The curriculum focused on managing one's anxiety and stress as well as effective ways to communicate with the patient. It was hypothesised that the curriculum, if efficacious, would directly lead to a reduction of anxiety and stress level of the caregiver. This reduction was hypothesised to produce additional benefits which would eventually leads to an improvement in **ED** related outcomes of the patient. These additional benefits formed the secondary focus of the **RCT** and two of the hypotheses were investigated in the following demonstration of the use of the new estimator.

The patients were recruited from a clinical unit specialising in **ED** in the Maudsley Hospital. The two routes by which the patients were admitted to the unit were via a visit to the emergency department following an acute medical event or on advice of the psychiatrist following an office visit. In order to be recruited for the study, the patient must have a formal diagnosis of **ED**, have a primary caregiver, and he or she as well as their caregiver must consent to taking part in the study.

5.4.1 The mediation hypotheses tested

Two mediation hypotheses are being tested. The first hypothesis (fig. 5.11) considers whether a reduction in caregiving stress as measured by the Family Questionnaire (**FQ**) led to an improvement in symptoms of eating disorders and consequently a reduction in the probability of relapse. The second hypothesis (fig. 5.12) similarly considers if a reduction of caregiving stress, also measured by the **FQ**, improves the anxiety levels of the caregiver, measured by the Accommodating and Enabling Scale for Eating Disorders (**AESED**), thus enabling the caregiver to better manage the symptoms of **ED**, resulting in a reduction in overall symptoms measured by the Eating Disorder Questionnaire (**EDE-Q**). The outcomes of interest in the two hypotheses also corresponds to a binary (relapse) and continuous (symptoms of **ED**) variable respectively. Both of these hypotheses had been suggested (Goddard et al., 2013; Magill et al., 2016) but not

formally tested. The hypotheses proposed, given that they had been suggested in the literature, are questions with genuine clinical implications and importance. However, the analyses presented here should not be interpreted for clinical significance since what is presented here deviates from the original analysis in some important ways to simplify the analysis in order to focus on the ability of the novel estimator to conduct sequential mediation for both continuous and binary outcomes of interest. Crucially, these hypotheses could not be tested by currently available estimators. The main deviation is in the handling of missing data where in the original analysis, multiple imputation was used but in this demonstration, only subjects with complete data were analysed. Next, I briefly introduce the demographics of the caregivers and patients.

Figure 5.11: CASIS example: Hypothesis one

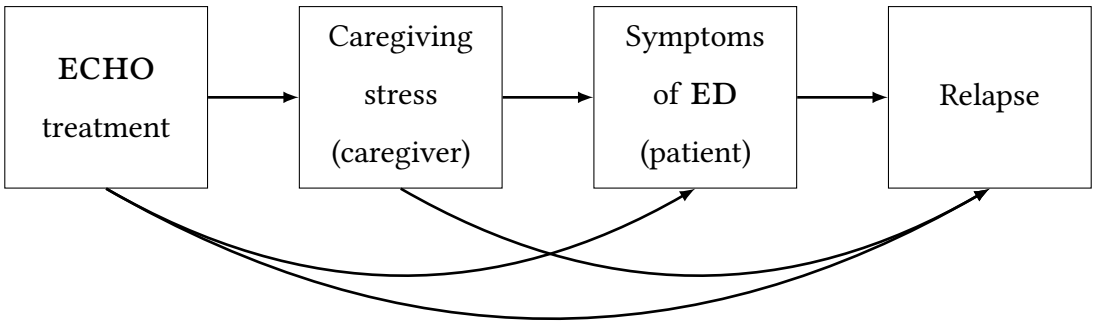
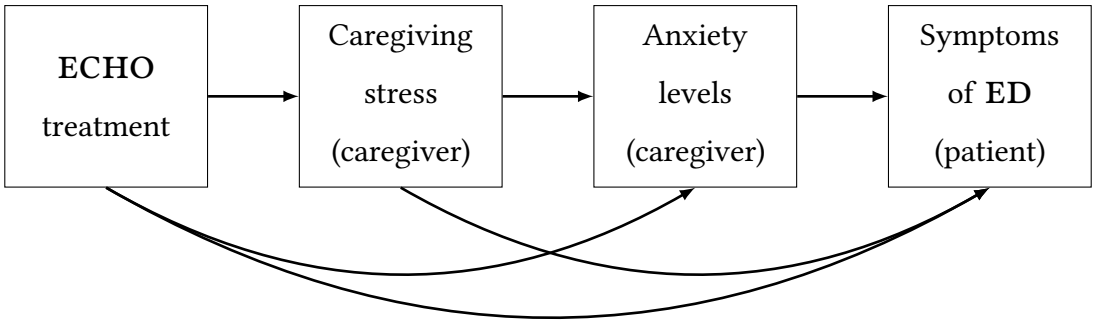


Figure 5.12: CASIS example: Hypothesis two



5.4.2 Descriptives of patients and their caregivers

The demographics of the patients and caregivers are summarised in table 5.1. Tables 5.2 and 5.3 summarises the variables used in the first and second hypotheses respectively. The caregivers and

patients were predominantly females who were white with an educational level of 'below degree'. Their mean ages were 51.73 and 26.99 respectively. Most of the caregivers are married while most of the patients are single. Most patients are on prescribed medication. The treatment and control group were generally similar in demographics and baseline measurements of mediators and outcomes with some differences in carers' gender, patients' education and patients' marital status.

Table 5.1: CASIS example: Sample characteristics (n=152)

	Caregivers						Patients					
	ECHO		Control		Total		**ECHO**		Control		Total	
Gender, Frequency (%)												
female	65	(85.53)	60	(78.95)	125	(82.24)	74	(97.37)	70	(92.11)	144	(94.74)
male	11	(14.47)	16	(21.05)	27	(17.76)	2	(2.63)	6	(7.89)	8	(5.26)
Age, Mean (SD)	50.92	(10.08)	52.55	(9.23)	51.73	(9.67)	26.26	(9.55)	27.73	(9.35)	26.99	(9.45)
Ethnicity, Frequency (%)												
white	71	(93.42)	72	(94.74)	143	(94.08)	71	(93.42)	72	(94.74)	143	(94.08)
asian	4	(5.26)	2	(2.63)	6	(3.95)	4	(5.26)	1	(1.32)	5	(3.29)
mixed	0	(0.00)	2	(2.63)	2	(1.32)	0	(0.00)	2	(2.63)	2	(1.32)
others	1	(1.32)	0	(0.00)	1	(0.66)	1	(1.32)	1	(1.32)	2	(1.32)
Education level, Frequency(%)												
above degree	4	(5.26)	11	(14.47)	15	(9.87)	3	(3.95)	6	(7.89)	9	(5.92)
degree	19	(25.00)	19	(25.00)	38	(25.00)	27	(35.53)	14	(18.42)	41	(26.97)
below degree	46	(60.53)	37	(48.68)	83	(54.61)	45	(59.21)	55	(72.37)	100	(65.79)
others	7	(9.21)	9	(11.84)	16	(10.53)	1	(1.32)	1	(1.32)	2	(1.32)
Marital status, Frequency(%)												
single	4	(5.26)	3	(3.95)	7	(4.61)	65	(85.53)	54	(71.05)	119	(78.29)
partnered	59	(77.63)	56	(73.68)	115	(75.66)	10	(13.16)	18	(23.68)	28	(18.42)
ex-partnered	13	(17.11)	17	(22.37)	30	(19.74)	1	(1.32)	4	(5.26)	5	(3.29)
On medication, Frequency(%)												
yes	NA	NA	NA	NA	NA	NA	57	(75.00)	54	(71.05)	111	(73.03)
no	NA	NA	NA	NA	NA	NA	19	(25.00)	22	(28.95)	41	(26.97)

Table 5.2: Mediators and outcomes for hypothesis one

Role	Variable	**ECHO**		Control		Total	
mediator 1, baseline	**FQ**, caregiver	2.46	(0.42)	2.41	(0.50)	2.44	(0.46)
mediator 1, 3 months	**FQ**, caregiver	2.39	(0.45)	2.44	(0.50)	2.41	(0.47)
mediator 2, baseline	**EDE-Q**, patient	4.32	(1.28)	4.15	(1.16)	4.23	(1.22)
mediator 2, 6 months	**EDE-Q**, patient	3.48	(1.60)	3.50	(1.47)	3.49	(1.53)
outcome, 12 months	Relapse, patient	45	(0.59)	45	(0.59)	45	(0.59)

* Numbers reported are mean (SD) except for relapse. Relapse is frequency (%).

Table 5.3: Mediators and outcomes for hypothesis two

Role	Variable	**ECHO**		Control		Total	
mediator 1, baseline	**FQ**, caregiver	2.46	(0.42)	2.41	(0.50)	2.44	(0.46)
mediator 2, baseline	**AESED**, caregiver	1.59	(0.72)	1.60	(0.78)	1.59	(0.75)
mediator 1, 3 months	**FQ**, caregiver	2.39	(0.45)	2.44	(0.50)	2.41	(0.47)
mediator 2, 6 months	**AESED**, caregiver	1.30	(0.75)	1.50	(0.80)	1.40	(0.78)
outcome, 12 months	**EDE-Q**, patient	3.16	(1.67)	3.46	(1.57)	3.31	(1.62)

* Numbers reported are mean (SD).

5.4.3 Results

The results of the first hypothesis are presented in table 5.4 and fig. 5.13. The **TE**, representing the overall effect of the Experienced Carers Helping Others (**ECHO**) treatment on the probability of relapse, was insignificant. The **DE** and all **IE**s were also insignificant.

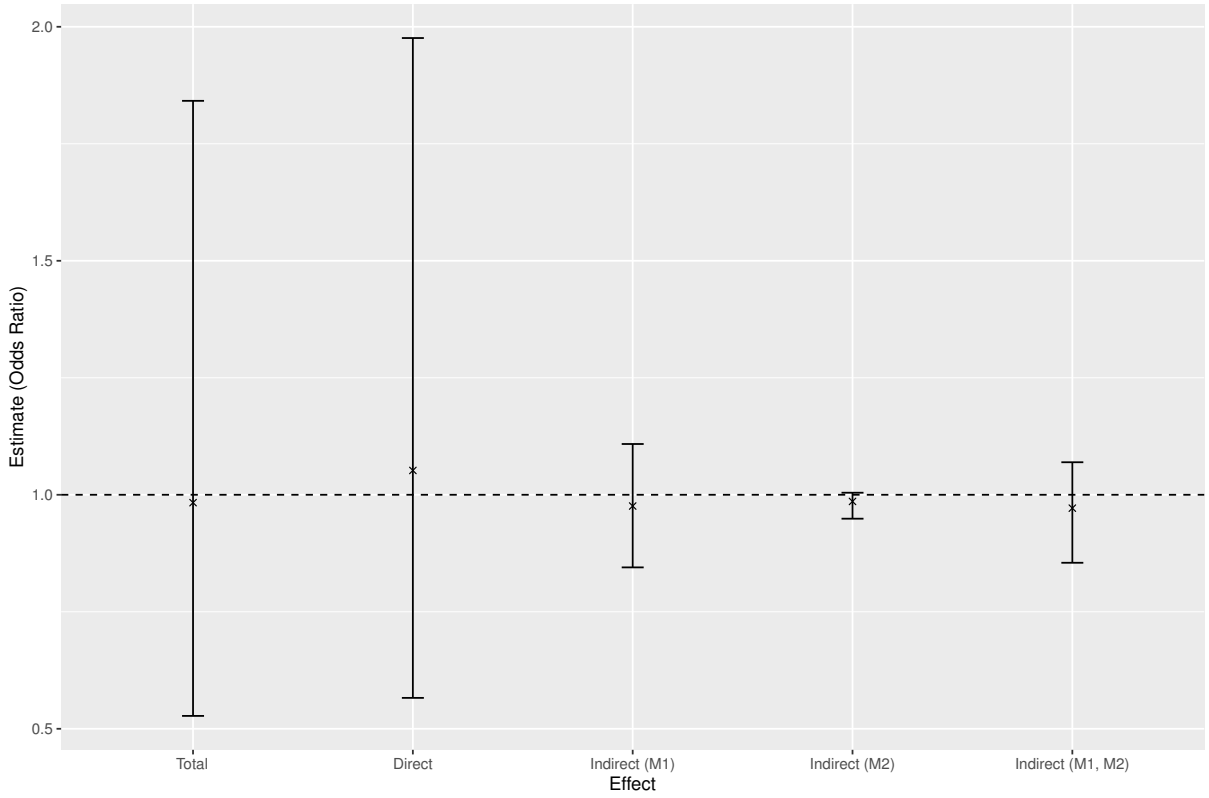
The lack of significance of the **TE** indicates that the data do not support the theory that the **ECHO** treatment reduces the chance of relapse. The lack of significance of the **DE** and **IE**s indicates that the data also does not support the theory that the **ECHO** treatment may act on the chance of relapse in direct and indirect pathways (reduction of caregiving stress of caregiver, reduction of **ED** symptoms of patient, or both). Notably, the indirect effect through *M2* (symptoms of **ED** of patient) borders on significance with the upper bound confidence interval at 1. This effect however is very small, at 0.99 suggesting that even if the effect had been significant, its magnitude is likely to be very small. The results for the second hypothesis

Table 5.4: Total and mediated effect estimates for hypothesis one

	est. (odds ratio)	95% CI
Total	0.98	(0.53, 1.84)
Direct	1.05	(0.57, 1.98)
Indirect (M1)	0.98	(0.84, 1.11)
Indirect (M2)	0.99	(0.95, 1.00)
Indirect (M1 & M2)	0.97	(0.85, 1.07)

are presented in table 5.5 and fig. 5.14. As was the case with the first hypothesis, the **TE**, representing the effect of the **ECHO** treatment on the symptoms of **ED** of the patient, was not significant. The **DE** and **IE**s were also not significant. The results of the **TE** indicates that there is a lack of support in the data for the theory that the **ECHO** treatment has an effect on the symptoms of **ED** of the patient. The results of the **DE** and **IE**s indicates that there is also a lack of support in the data for the theory that the **ECHO** treatment has an effect on the symptoms of **ED** of the patient via direct and indirect pathways (caregiving stress of caregiver, anxiety levels of caregiver, or both). Notably the **IE**s all border on significance with each of their upper bound confidence intervals very close to 0. The magnitude of each of the **IE**s, -0.07, -0.03,

Figure 5.13: Graph of total and mediated effect estimates for hypothesis one

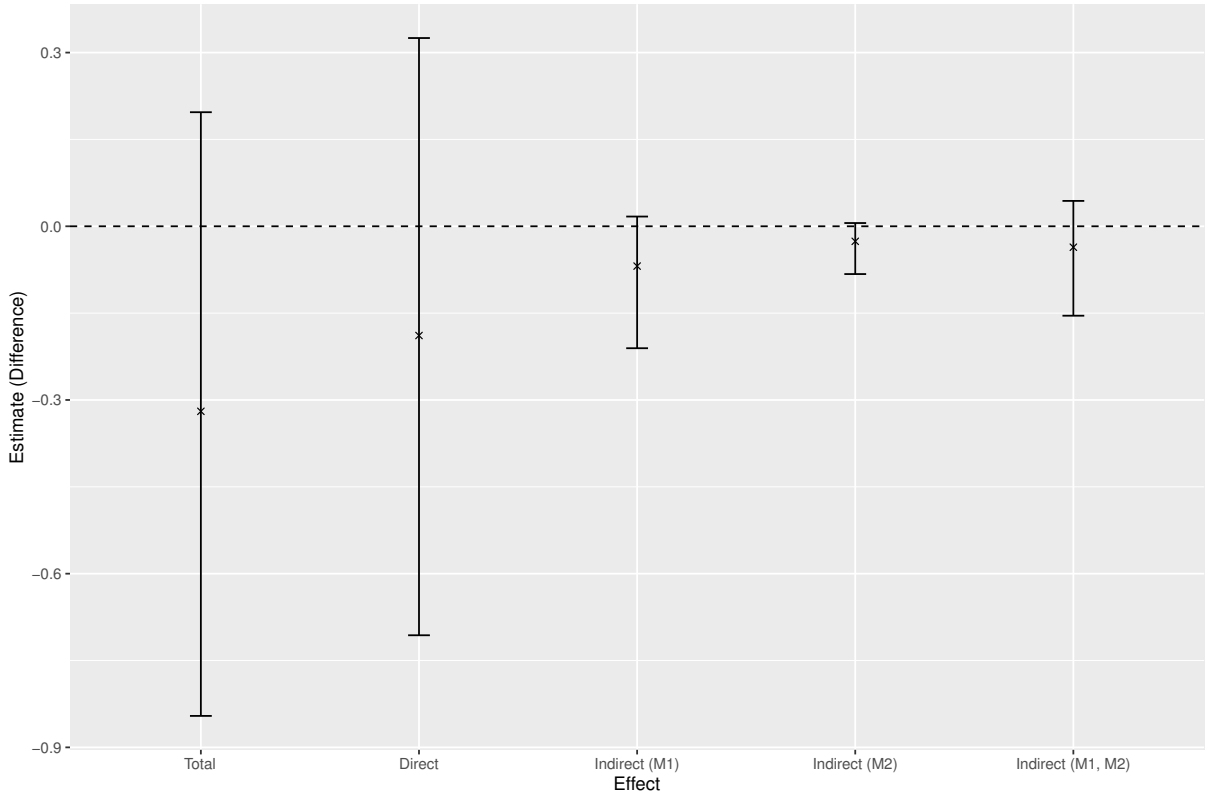


and −0.04 are also very small when compared to the mean of the outcome (**EDE-Q**, mean = 3.31) at baseline indicating that even if the **IE**s were significant, the effect are likely to be small. Whether or not these small effects are clinically important is a judgement that the clinician has to make.

Table 5.5: Total and mediated effect estimates for hypothesis two

	est.	95% CI
Total	-0.32	(-0.85, 0.20)
Direct	-0.19	(-0.71, 0.32)
Indirect (M1)	-0.07	(-0.21, 0.02)
Indirect (M2)	-0.03	(-0.08, 0.01)
Indirect (M1 & M2)	-0.04	(-0.15, 0.04)

Figure 5.14: Graph of total and mediated effect estimates for hypothesis two



5.4.4 Discussion of the results

The results provided a test of whether or not the hypothesised theories were supported by the data collected. Crucially, we were able to test path-specific hypothesises and to consider whether or not the evidence supported each pathway of interest. While the current results suggest that the proposed theories are not supported by evidence, it is worthy to note that some of the indirect pathways border on significance even though the **TE** is clearly not. This suggest that such mediational analysis can provide unique insight into possible intermediary pathways when the **TE** is not significant especially when the focus had been on the **TE** in previous analyses of similar treatments. Intermediary pathways that are significant may represent previously pathways of mechanism that were not within theoretical consideration and provides a strong indication for these pathways to be further investigated. Beyond that, mediational analysis can also aid in identifying which pathways are supported and which are not, leading to a revision of the theoretical model.

Contextualising the role of mediated analyses to the results from **CASIS** where all the effects

were not significant, we can use the results to update the theoretical model by ruling out the mediated pathways as mechanisms of change. While this does not seem to be a noteworthy result, the importance of the results lie in challenging our preconceived notions of how change occurs in the treatment of patients with **ED**. The lack of significance for the mediated effects does not mean that mediated effects do not exist. They could still exist via other pathways and any consideration of alternative mechanisms will need to be informed by both existing literature and expert knowledge of the clinician. To be clear, these additional considerations do not change the fact that there was a lack of overall efficacy of the treatment on the outcomes of interest. Should mediated pathways be found and substantiated with evidence from a mediation analysis, this information will improve our understanding the processes that underpin changes in the patient with **ED** or their caregiver. This improved understanding can potentially translate into revisions to the treatment protocol to make it efficacious.

Notably, this application of mediation analysis on **CASIS** is novel in two ways. Firstly, unlike existing methods, this application of mediation analysis tests a hypothesis which defines the causal effect as an **OR** (hypothesis one). Secondly, the hypotheses tested involved two mediators, one after another sequentially. Both analyses presented showed no significant results, but one main value of such analyses is to allow the testing of theoretical mechanisms of action. In this case, the treatment is unlikely to have acted through the pathways tested. This enables one to revise and consider alternative mechanisms of actions and eventually update the theoretical model once a mechanism of action has been identified. The analysis shown in this chapter however has some limitations since it was conducted under the assumption that there was no unmeasured confounding between mediators and outcomes. This assumed there were no hidden variables that drove both the mediators and the outcome. In the two mediators case, this is recasted as an assumption that there was no confounding between mediators and between mediators and outcome. This is an assumption because there is no way to provide evidence that there is no unmeasured confounding, i.e. to prove a negative. What is possible however is to assume that there was indeed confounding and assess the effect that this confounding might have on the estimated mediated causal effects. If this effect was deemed to be small, then the results are trustworthy. If the effect was deemed to be large, then a judgement needs to be made on how likely this confounding took place and if it took place, how large is the magnitude of confounding and in which direction. These additional analyses are known as the sensitivity analysis and are discussed in the next chapter.

5.5 Discussion

The chapter first laid out a series of methods used to validate the simulation-based estimator and aimed to answer the question of 'Can we trust this novel estimator?'. The validation procedures were conducted across different samples and effect sizes as defined by the Cohen's D. An adaption of the Cohen's D to cater to binary outcomes was developed to allow a consistent definition of effect size across the validation checks. This adaptation built on the idea of using a difference of probabilities as the causal estimand to construct a Cohen's D. This adaptation enabled comparability between the checks across the different model combinations.

The two parts of the study investigated the magnitude of precision loss due to a larger number of mathematical operations and the discrepancies between the model- and simulation based estimates. For the model combinations with a model-based estimate for mediated causal effects, the two sets of estimates between the model- and simulation-based estimates were compared and found to be nearly identical with the exception of the model combination *BBB*. The differences were small particularly when considered within the width of the respective confidence intervals and are most likely due to precision loss in line with the observations in the sub-study on precision loss. For the model combinations without a model-based estimate for mediated causal effects, only the **TE** estimates were compared. This uses a property of the **TE** which can be estimated using two different methods. This part of the study rests upon the idea that the simulation-based estimator uses the same procedures for conducting the estimation, regardless of which effect it is being used to estimate. The only difference between the parameters used for the estimation between different effects are the values of the treatment indicator. These values of the treatment indicator represent the different **PO** and are derived from the definition of the mediated causal effects discussed at length in chapter 4. The idea is that since the procedures are the same, validating the procedures by using them to estimate **TE** and comparing that with the model-based estimate, we can validate the model combinations for which model-based estimators cannot estimate the mediated causal effects.

The results provide a clear indication that the estimates between the model- and simulation-based estimates are nearly identical for continuous outcomes. The differences observed for binary outcomes are small when considered in their respective confidence intervals and are also what

was to be expected as a result of precision loss due to larger number of mathematical operations. The same precision loss is observed for model-based estimates as discussed in the sub-study for precision loss.

The later part of the chapter focused on how this novel estimator is able to estimate causal effects in an **RCT**. The aim of this application is to demonstrate the real world applicability of this estimator. The results of the application indicate that previously considered hypotheses are not supported by the data from this trial. The analysis done in this chapter with a binary outcome may have been possible with existing methods but these methods each have significant limitations (e.g. methods by Valeri and VanderWeele (2013)) or uncertainty about the impact of choice of parameters used in the estimation (e.g. the choice of integration points in Mplus). These limitations and uncertainties were previously discussed. The work conducted in this thesis overcame these limitations and also provided a set of procedures to verify the results from the novel estimators. Furthermore, the analyses could accommodate two mediators in sequential order, allowing investigation of specific part of a theoretical model. Both the ability to conduct causal mediation with binary outcome and causal mediation with two mediators represent the improvements over current methods that this estimator brings about.

More importantly, the methods developed in this thesis removed a significant limitation in applying mediational analyses. This removes the need on the part of the researcher to consider whether or not a hypothesis can be analysed for mediated causal effects and shifts the focus towards considering what potential mediated causal relationships to test for. This is useful for informing a theoretical model since these analyses allow testing of hypotheses that coincide with proposed pathways of change indicated by the theoretical model. These analyses also provide insight into mechanisms of change by considering mediated causal pathways, which can exist even if the **TE** is not significant. These mediated causal pathways can also be examined in and of themselves to start to disentangle why a treatment worked or did not work.

In summary, this chapter provided the evidence that the simulation-based estimator is similar in terms of characteristics to the model-based estimator and demonstrated what the novel improvements in this estimator allows one to do in an **RCT**.

Chapter 6

Sensitivity analysis for novel causal effect estimator

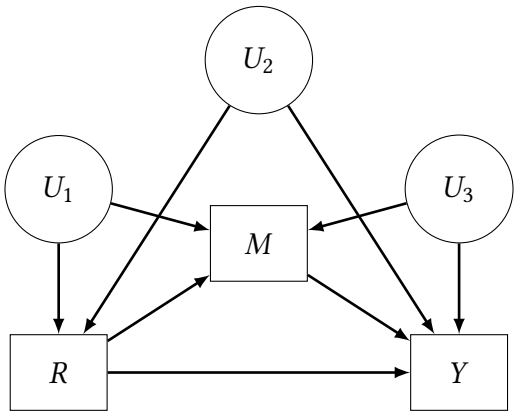
6.1 Introduction

The last chapter ended with a discussion of the procedures used for validating the causal effect estimates as well as how the causal effect estimator can be used in a real life scenario. In this chapter, I will discuss one remaining issue in the use of the causal effect estimator, namely, the effect of violation of assumptions on the causal effect estimates. As discussed in chapter 4, the causal effects were estimated with three important assumptions:

1. No treatment-mediator confounding
2. No treatment-outcome confounding
3. No mediator-outcome confounding

The three assumptions are each represented by an absence of U_1, U_2 and U_3 in fig. 6.1 where R, M and Y represents the treatment, mediator and outcome respectively. The direction of the arrows represent the direction of causation. The aforementioned assumptions remain as assumptions because as stated, they are statements about the non-existence of a relationship. To fully prove non-existence would require that all possible confounders be enumerated, quantified and the respective data collected on them. They could then be tested and shown not to be

Figure 6.1: Mediated effects with confounding



a confounder. This is impractical since possible confounders will include both known and unknown confounders. For known confounders, they should ideally be addressed by collecting information on them as part of the trial. For unknown confounders, quantifying them would then imply that data on all possible variables be collected since any variable could potentially be an unknown confounder. This is impractical since any such attempts will unlikely be exhaustive. An alternative strategy therefore needs to be used to assess the impact of unknown confounders on the causal effect estimates. The *sensitivity analysis* is one such strategy that can be used to assess the effect of different levels of confounding on these estimates.

The first and second assumptions are violated when there is a common cause of the treatment with the mediator or outcome respectively. This is unlikely to occur in an randomised controlled trial (**RCT**) since the treatment offered is randomised and thus cannot be predicted. Since these two assumptions are unlikely to be violated, the assumptions are considered to be justified in an **RCT** and focus will be placed on the third assumption of no mediator-outcome confounding. If the treatment was not randomised for example in observational studies, then the first two assumptions will need to be justified either through the use of a sensitivity analysis or by other means.

The sensitivity analysis, as the name suggests, assess how *sensitive* the causal estimates are to a violation of assumptions. The analysis seeks to answer the question: What would the causal estimates be if a confounder of M and Y is present? Breaking this question down, we have the following:

1. At what level of confounding will the causal effect estimate of interest be rendered null or

insignificant?

2. In what direction does increasing levels of confounding change the causal effect estimate of interest?

Level in this context refers to a measure of the strength of the confounding, i.e. the extent of violation of the assumptions. The stronger the relationship, the more severe is the violation of the assumptions. Quantifying this strength is addressed in more detail later in this section. The analysis entails simulating scenarios where the assumptions are violated and the corresponding causal effects are estimated. This simulation is repeated for varying degrees of violations. If the causal effect estimates change in a meaningful way under even minor violations of the assumptions, a greater scrutiny of the likelihood in which the assumptions are violated in reality is then warranted. On the other hand, if the causal effect estimates do not change meaningfully under severe violations of the assumptions, the conclusions reached in the unconfounded analysis then warrant a higher level of confidence than if the estimates were changed meaningfully. The implemented sensitivity analysis follows closely to that implemented in the *R* package *mediation* (Imai et al., 2010) with some important novel additions I have made.

The *mediation* package uses *seemingly unrelated regressions* while the sensitivity analysis I implemented uses *latent variable models*. The use of the seemingly unrelated regression is not stated in the documentation of the package but is found in the comments in the source code of the software. A study of the code involved also confirmed that the implemented method for sensitivity analysis in *mediation* is seemingly unrelated regressions. The use of latent variable models instead, as in this thesis, enable the implementation of sensitivity analysis for logistic regressions. This is a novel development not currently implemented in any existing software for the estimation of mediated causal effects. The use of a latent variable model for sensitivity analysis additionally has the advantage of using the same conceptual framework for both normal and logistic generalised linear model (**GLM**).

This sensitivity analysis introduces a number of concepts that need to be defined and quantified, namely, the degree of violation of assumptions, the meaningfulness of change, and the likelihood in which the violations are present in the context of the trial. The degree of violation will be addressed in the next section, while I will address the latter two here. The meaningfulness of change refers to a change in the causal estimate under different degrees of violation of the

assumptions. The concept of what is a meaningful change cannot be ascertained by statistical means and needs to be addressed with subject matter knowledge of the trial in question. This is the same for the likelihood in which the violations are present. This requires consideration of potential variables that could act as confounders and determine how likely they are to be acting as confounders in the context of the trial. A failure to consider the likelihood of these confounders directly threatens both the internal and external validity of the results of the trial. One strategy to address both the meaningfulness of change and the likelihood of violation is to refer to prior literature to quantify both aspects. This however needs to be done carefully to ensure similarity in context between the literature and the trial in question.

This chapter will next discuss how confounding is quantified and simulated first for a single mediator and subsequently for two mediators for normal **GLM** in order to conduct sensitivity analysis for the models discussed in the thesis. The discussion next focuses on the case of the logistic regression and how the confounding is conceptualised. Special attention will be placed on the differences between the logistic regression where the probability of an outcome occurring is not directly observable and the case of the normal **GLM** where the outcome is directly observable. Finally, going back to the example Carers' Assessment, Skills and Information Sharing (**CASIS**) trial, the sensitivity analysis was applied to determine the effects of confounding on the resulting causal estimates as a demonstration of the utility of the sensitivity analysis procedure.

6.2 Assessing effects of confounding

The first step to assessing the effects of confounding is to first quantify the relationship between the confounder with the mediator and outcome. I will focus first on the single mediator case before discussing the two mediator case. The discussion here will be restricted to the normal **GLM**. The case of the logistic regression will be discussed in a later section.

Referring to fig. 6.2, the violation of the assumption of interest is represented by the presence of U. Conceptually, the assumption states that there is no confounding between the mediator and the outcome. This conceptualisation of confounding makes use of a latent variable, U with a causal relationship between the error terms of both the mediator and outcome. The same form of confounding can alternatively be conceptualised as a correlation between the two error

terms as shown in fig. 6.3. In both figures, R, M and Y represent the treatment, mediator and outcome respectively. ϵ represents error terms, single-headed arrows represent the direction of causality and double-headed arrows represent correlation. Both forms of conceptualisation of confounding are equivalent in what they represent and their differences lie in how they are each parametrised. Given the equivalency, the parameters from one model can be converted to the parameters of the other model and vice versa. This thesis focuses on the use of the latent variable conceptualisation and parametrisation, the reasons for which are discussed in a subsequent section on the logistic model. One difference between the U of fig. 6.2 and U_3 of fig. 6.1 are the arrows that point away from U and U_3 as well as the presence of ϵ_3 and ϵ_2. The latent variable in fig. 6.2 points at the error terms of the mediator and outcome variables rather than the variables themselves. This difference in parametrisation of the model used for sensitivity analysis first considers the shared variance between the outcome and the independent variables and covariates. The unaccounted-for variance, represented by the variance of the error term can then be conceptualised as having a shared correlation with the error term of another model. In the case of a single mediator, this would be a shared correlation between the error terms of the mediator and outcome model. This shared correlation can then be alternative parametrised as the error terms sharing a correlation with a third latent variable that is randomly generated to be correlated with both the error terms. This second parametrisation is the latent variable parametrisation used in this thesis. The latent variable is parametrised to be drawn from a standard normal distribution and with this parametrisation, the model for sensitivity analysis becomes over identified rather than under-identified. This way in which the confounding variables have been conceptualised is similar to how factor analysis using structural equation model (**SEM**) is parametrised, thus software used to estimate **SEM** models are used to conduct the sensitivity analysis. The conversion between the latent variable conceptualisation and the

Figure 6.2: Mediation with Y, M confounding: latent variable formulation

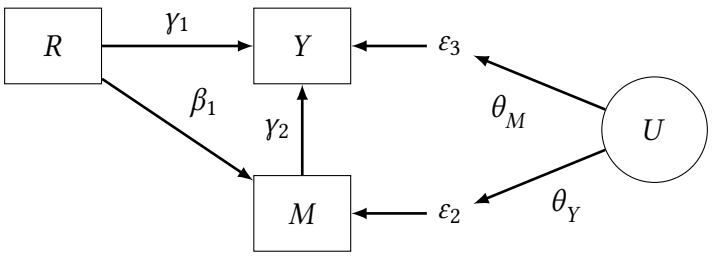
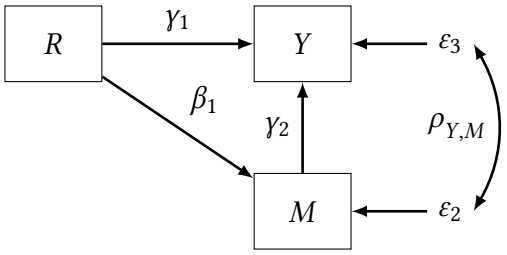


Figure 6.3: Mediation with Y, M confounding: correlated error terms formulation



correlation conceptualisation, is as follows:

$$\begin{aligned} Cov(\varepsilon_3, \varepsilon_2) &= \theta_M \theta_Y \\ \rho_{Y,M} &= \frac{Cov(\varepsilon_3, \varepsilon_2)}{Var(\varepsilon_3)Var(\varepsilon_2)} \\ &= \frac{\theta_M \theta_Y}{Var(\varepsilon_3)Var(\varepsilon_2)} \end{aligned}$$

By setting different values of θ_M and θ_Y, we can simulate different degrees of confounding. The choice of what values to use for setting the θs rests upon the idea that the θs can be converted to a correlation parameter and vice versa. Since a correlation is bounded between 0 and 1, we can then select equally spaced level of confounding using correlation and then converting this back to the θs. For a given set of M and Y models for which we want to conduct sensitivity analysis on, it needs to be noted that there is a maximum degree correlation between the error terms of M and Y and hence implying that there are limitations in the values that the θs can assume. Think of this as having three variables, A, B and C. If the correlation between A and B, and A and C are known, then the correlation between B and C has to fall within a certain range because it is now constrained by the other pairwise correlations. Likewise in the case of setting the values of θs, they are constrained by the maximum correlation that can be induced between the ϵs. This value of a maximum correlation was determined in this thesis through a set of optimisation procedures which quickly tests out the use of different values of correlation to determine the point at which the software fails to estimate. *Mplus* was used to test out these different values of correlation and when the maximum possible correlation is exceeded, *Mplus* produces error messages to the effect that the model cannot be estimated. The optimisation procedure and the procedure to identify the maximum correlation has been implemented in *R* which then interfaces with *Mplus*. The optimisation procedure additionally allows one to specify the level of precision required for the maximum correlation and is set at a default of three decimal places. At the end of deriving the maximum correlation, equally spaced correlations

are then selected to form the degrees of violations to be tested. These correlations are finally converted back to θs for use in simulating the effect of confounding.

Lastly, the parametrisation using correlation and latent variables differ in one additional way. In the latent variable formulation, we can change the path coefficient of U on the mediator and outcome independently while in the correlation parametrisation, it assumes that the path coefficients are proportional to the variance of each of the ε for M and Y. This proportional relation is used in the conversion between the correlation to the path coefficients and it also implies that the degree of confounding is proportional to the variance of the error terms. This is justifiable because if the variance of the error terms are large, it implies that a large proportion of variance of the dependent variable is unaccounted which makes it more likely that a confounder is present to account for this variance.

Once the set of different degrees of confounding is obtained, the estimation of the causal estimates under confounded conditions proceed using the same estimating procedures as the mediated causal effects discussed in chapter 4. This is repeated for each level of confounding and for all the effects of interests. The next section expands upon the use of sensitivity analysis to accommodate more than a single mediator.

6.3 Sensitivity analysis for two mediators with normal GLM

The sensitivity analysis for two mediators compared to a single mediator has an extra complication. Since there are two mediators, the assumption of no mediator-outcome confounding means that there is the added consideration of the extra mediator. Instead of two ϵs, there are now three. The first step is to identify and choose a method to parametrise the sensitivity analysis for two mediators. Firstly, with two mediators, the assumption of no mediator-outcome confounding will be modified to no mediator-mediator and no mediator-outcome confounding. This is because the second mediator acts as an outcome of the first mediator and thus the assumption of no mediator-outcome confounding also implies no mediator-mediator confounding when there is more than a single mediator. The first mediator will be referred to as M_1 and the second mediator as M_2. With the added no mediator-mediator confounding, for a causal model with two mediators, we thus have three pairwise confounding relationships to consider: M_1 with M_2,

M_1 with Y, and M_2 with Y. Continuing to work with latent variables, there are two ways in which the confounding can be parametrised. The first is to assume that all three variable are confounded by the same variable and the second way is for each pairwise confounder to be independent. These two parametrisations are represented by fig. 6.4 and fig. 6.5 respectively.

Figure 6.4: Mediation with Y, M_2, M_1 confounding: single latent variable formulation

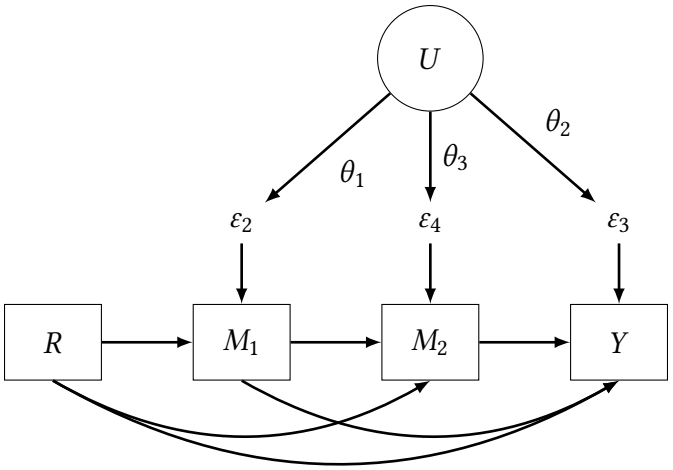
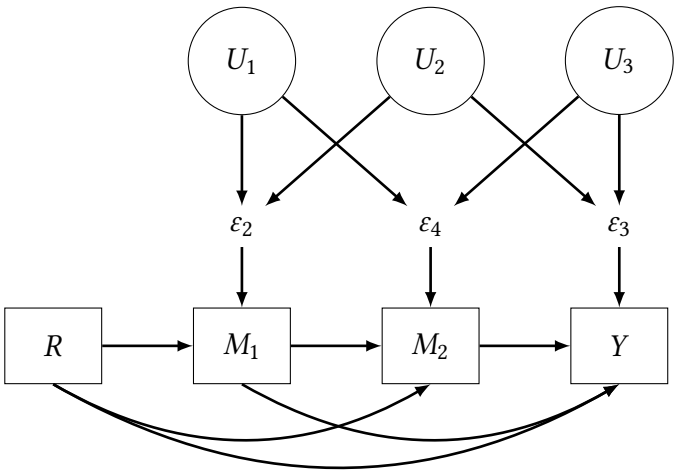


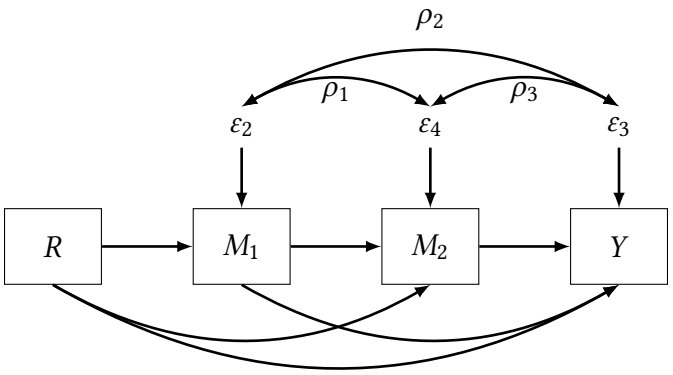
Figure 6.5: Mediation with Y, M_2, M_1 confounding: latent variable formulation



The first parametrisation where there is only a single confounder is unlikely since there is no requirement or restriction for the mediators and outcomes to be related in any specific way. The second parametrisation is thus more likely and in the event that the first parametrisation is the correct model of confounding, the second parametrisation will simply reproduce what happens in the first parametrisation with very closely correlated latent Us. The corresponding model for the second parametrisation using correlation is represented by fig. 6.6. The correlation

parametrisation is shown mainly to demonstrate that in the case of normal **GLM** models, both parametrisations exist and are equivalent. Similar to the one mediator case, a similar relation-

Figure 6.6: Mediation with Y, M_1, M_2 confounding: correlated error terms formulation



ship exists between the ρs and the Us. Using ρ_1 and U_1 as an example we have the following relationship.

$$\begin{aligned} Cov(\varepsilon_2, \varepsilon_4) &= \theta_{U_1,M_1}\theta_{U_1,M_2} \\ \rho_1 &= \frac{Cov(\varepsilon_2, \varepsilon_4)}{Var(\varepsilon_2)Var(\varepsilon_4)} \\ &= \frac{\theta_{U_1,M_1}\theta_{U_1,M_2}}{Var(\varepsilon_2)Var(\varepsilon_4)} \end{aligned}$$

The path $U_1 - M_1$ will be notated as θ_{U_1,M_1} and the path $U_1 - M_2$ will be notated as θ_{U_1,M_2}. As with the one mediator case, the conversion between correlation with the path coefficients assumes that the strength of confounding is proportional to the variance of the error terms of the respective pairs of dependent variables.

I have thus far discussed how the sensitivity analysis proceeds with continuous dependent variables modelled using the normal **GLM**. In the next section, I will discuss sensitivity analysis with binary dependent variables modelled using logistic regression.

6.4 Sensitivity analysis for one mediator with logistic GLM

Binary dependent variables in this thesis are modelled using logistic regressions and one important difference between the logistic **GLM** and the normal **GLM** is the difference in interpretation of the error term in the logistic **GLM**. Note that in some literature, it is stated that there are

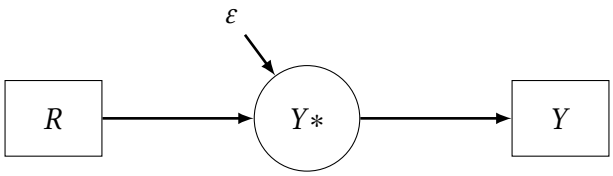
no error terms in a logistic **GLM**. This is due to the idea that in a logistic **GLM**, there are no residuals because rather than model the observed dependent variable which is binary, the probability of the dependent variable is modelled instead. Since the observed dependent variable is not modelled directly as in the normal **GLM**, there are no estimands of any error terms. Additionally, the distribution of a binary variable is parametrised using only a single parameter, its mean. The variance is derived from the mean. This also meant that the mean and variance cannot be modelled independently thus the concept of the error term as a representation of unaccounted variance as in the normal **GLM** case is not possible.

An error term in the logistic regression however can exist for two purposes. The first is as a representation of the unaccounted variance but this use of the error term is different from that of the normal **GLM**. The error term is always drawn from a standard logistic distribution and the values of the regression coefficients of a logistic regression is sized according to the error term. This meant that the ratio of variances between the untransformed by the inverse link function of a predicted value of a logistic regression with the error term provides an indication of the amount of variance that had been accounted for. This is because the distribution of the error term never changes and if the ratio is large, it means that the variance of the predicted value is much larger than that of the error term and vice versa. The function of the error term is thus to provide scale to the predicted values. The second use of the error term is to realise the dependent variable. Without the error term, we can realise the dependent variable by using the predicted probability of the dependent variable. Alternatively, we can make a draw from the logistic distribution and add this value to the untransformed predicted value of the dependent variable. If this value exceeds one, the realised binary variable is one and if below one, it would be zero. This is equivalent to realising the dependent variable using predicted probabilities. The drawing of the error term from a standard logistic distribution is thus simulating the fundamental uncertainty discussed in chapter 4.

The sensitivity analysis of a binary outcome will make use of both purposes of the error term. The conceptualisation of confounding using the formulation of a latent U as a cause for the error terms thus takes on a slightly different meaning as seen in fig. 6.7. Since the variance of the error terms is now relative to the predicted values of the logistic regression, a U that can predict this error term implies a U that can predict the fundamental uncertainty of the model. This indicates that a least part of the fundamental uncertainty is not uncertain at all since it can be

predicted by U. Furthermore, since the variance of the error term is proportional to the variance of the predicted value of the logistic regression, the maximum amount of confounding that can occur is the variance of the error term. For a dependent variable with very little unexplained variance, the variance of the predicted values will be much larger than that of the error term and thus very little confounding in the form of U predicting the error terms can take place. One complication of this formulation is that since the error term exists to provide scale and its variance does not change, there is no limit on the maximum amount of correlation between error terms which exists in the normal **GLM**. Thus when the mediator and outcome models consist solely of binary dependent variables, there is no need to test for the maximum possible amount of correlation. A maximum amount of correlation can exist when there is a mixture

Figure 6.7: Logistic regression: latent variable formulation



of logistic and normal **GLM** models and the same procedure for determining this maximum amount is conducted as what was done for the normal **GLM** case.

6.5 Implementation details

The aforementioned methods for conducting sensitivity analysis were implemented in *R* and *Mplus*. *Mplus* was used specifically for its ability to conduct latent variable modelling with logistic regression.

1. Identify the maximum amount of correlation possible. This step is skipped if all the mediators and outcome are binary.
2. Generate the different degrees of confounding to test for between 0 and the maximum amount of confounding.
3. Estimate the model with the constraints on the path coefficients from the latent confounders on the mediators and outcomes.
4. Use the newly estimated path coefficients and variance-covariance matrix of the path

coefficients to estimate the resulting causal effects following the procedures laid out in chapter 4.

5. Present the results graphically for comparisons with the original causal effect estimate.

One important note on the implementation of the sensitivity analysis is that the model used for 0 confounding is not the same as the original causal estimate with the assumption of no confounding. This is because in the sensitivity analysis, 0 confounding is modelled by constraints on the path coefficients from U to the mediators and outcomes (set to 0). These constraints meant that there was more certainty in the parameters of the model which is reflected in the much smaller confidence intervals of a model with 0 confounding in the sensitivity analysis and the original causal effect estimate. This is demonstrated in the subsequent section demonstrating the use of the sensitivity analysis on CASIS.

6.6 Application with CASIS

The CASIS trial, used previously to demonstrate the estimation of mediated causal effect is used again to demonstrate the conduct of a sensitivity analysis. For each hypothesis, a separate sensitivity analysis needs to be conducted. figs. 6.8 and 6.9 illustrate the results of each analysis. From the results, we first note that as indicated previously, the confidence intervals of the model with 0 confounding are smaller than the original causal effect estimates for all the effects of interest. Next we note that the confidence interval of each effect gets wider with increasing confounding and this is to be expected since a larger amount of confounding meant that the independent variables had a smaller amount of shared variance with the dependent variable. In the case of the CASIS, all the estimates across different levels of potential confounding remain non-significant suggesting that with or without confounding, the causal effects are unlikely to exist. Should some of the original causal effect estimates be statistically significant, this analysis could then be used to identify the degree of confounding that would render such an estimate statistically non-significant. In this situation the idea would then be to make a judgement about how likely the stated amount of confounding exists and therefore how likely the confounding was to have the effect found using this sensitivity analysis.

Figure 6.8: Graph of sensitivity analysis for hypothesis one

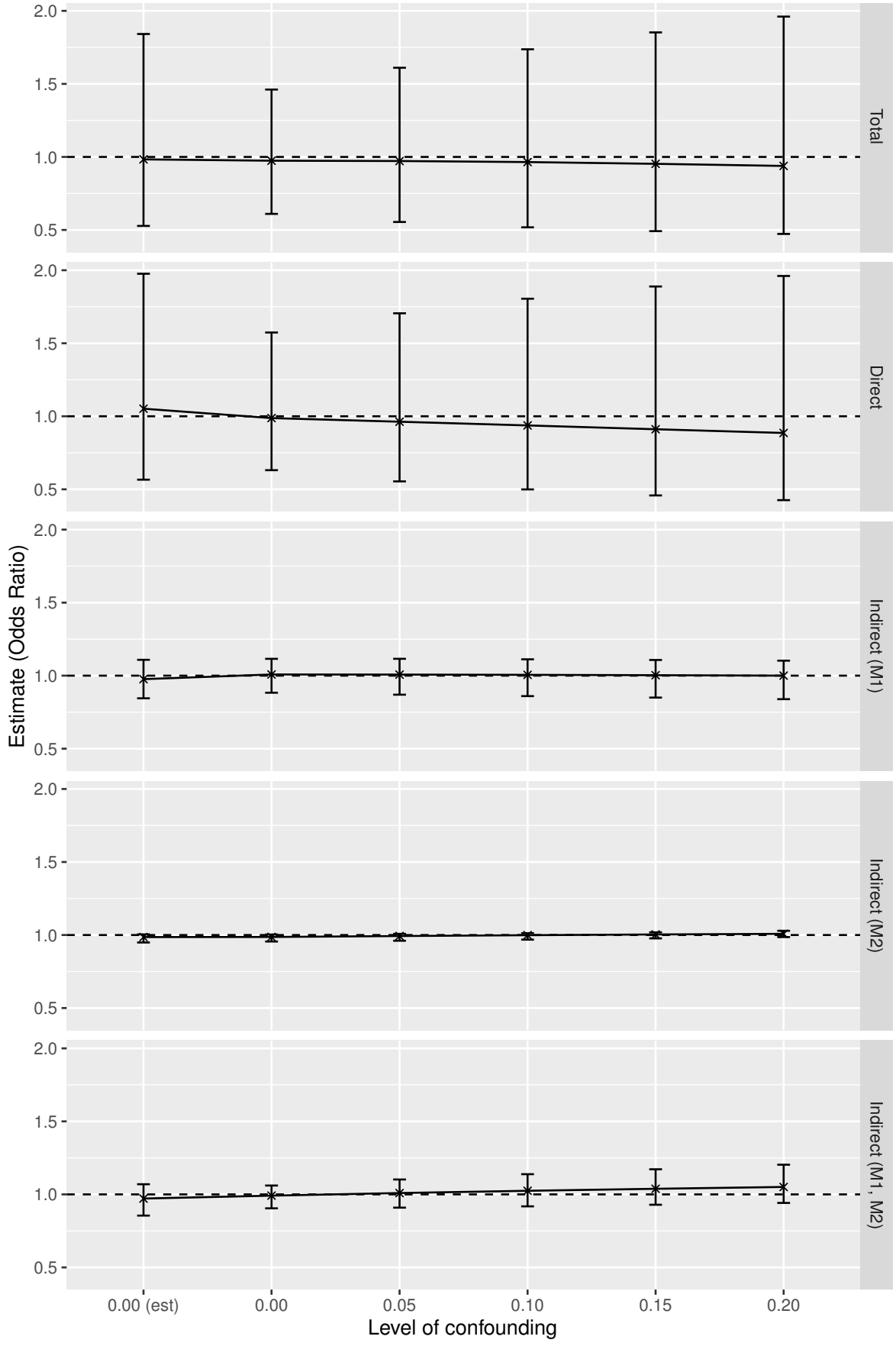
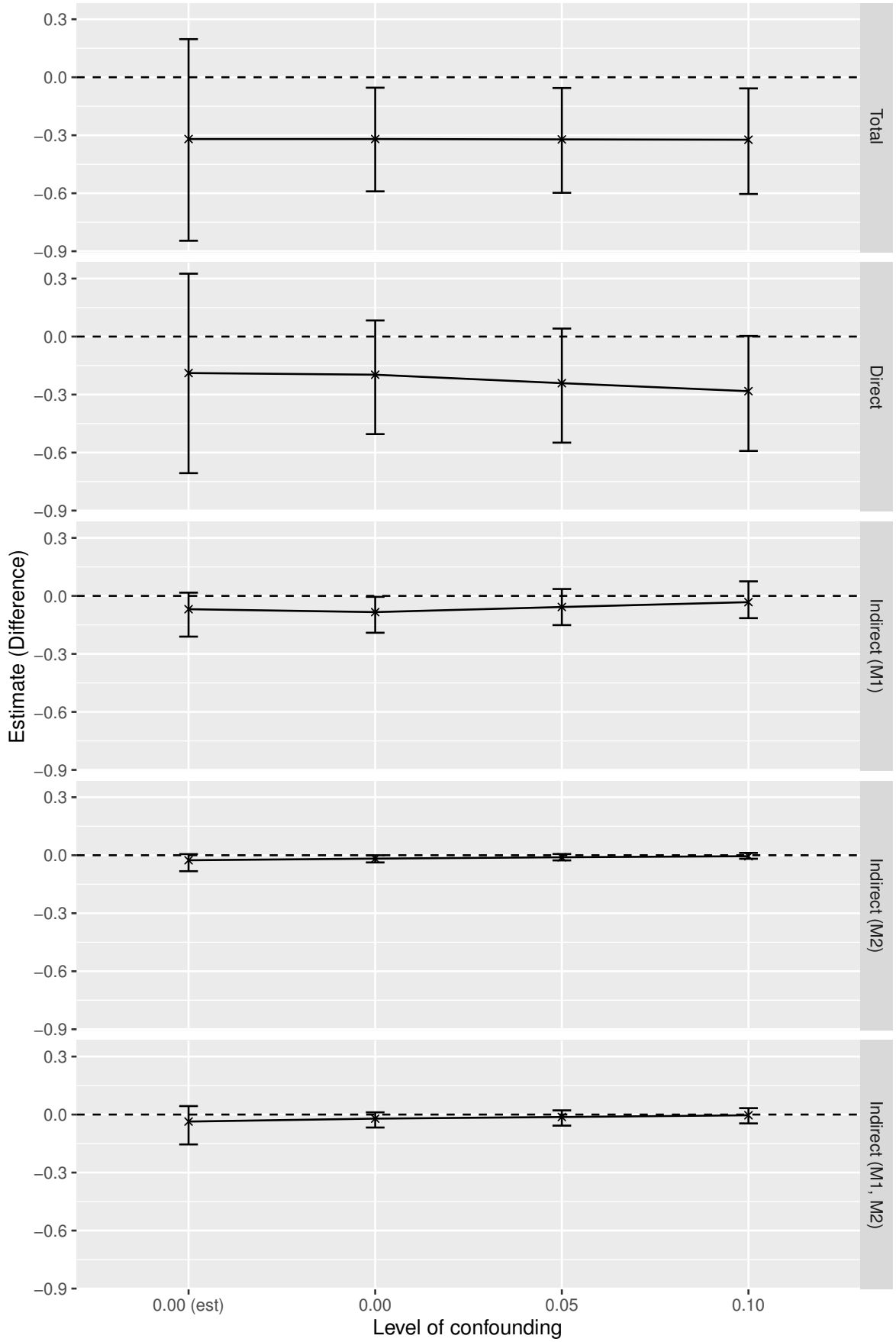


Figure 6.9: Graph of sensitivity analysis for hypothesis two



6.7 Discussion

This chapter discusses the need and purpose of sensitivity analysis in the estimation of mediated causal effects. The need for the sensitivity analysis arose from assumptions that were made, specifically in the case of an **RCT**, the absence of mediator-outcome confounding. Sensitivity analyses were required to ascertain how the causal effect estimates would change should the assumptions be violated since ruling out the presence of confounders is unlikely to be practical. This set of analyses provide an estimate of the direction and magnitude of potential bias under varying degrees of violation of the assumption. With this set of analyses, should the results suggest that the causal effect estimates are likely to go from being significant with the assumptions to insignificant with a violation of assumptions, it is then up to the researcher to justify the assumptions and to consider how likely the said degree of violation exists. This can be done through a review of prior literature and cannot be determined using any statistical methods.

In the next chapter, I discuss some of the significant contributions from this thesis and broader implications of the work done.

Chapter 7

Discussion: Revisiting aims, significant contributions and concluding remarks

In this concluding chapter of the thesis, I will be briefly revisiting how the aims of this thesis came about and the steps taken to achieve them, highlighting the important developments that this piece of work is built upon. Next, I will discuss the extent to which the aims had been met, focusing on the strengths and weaknesses of the methods developed. I will also highlight how some of the methodological developments have wider applicability than the methods presented in this thesis. These developments are significant because they can serve as building blocks for the development of a general method for the estimation of mediated causal effects parametrically. Lastly, I would end this chapter with avenues of further developments of the methods in this thesis.

7.1 Retracing motivations and methods development

The primary question that this thesis sets out to answer is 'how can we use methods of causal inference in process evaluation (**PE**) to answer questions of *how* or *why* an intervention works?'. The question had within it an assumption that methods of causal inference are not widely used in the conduct of **PE** so the first task was to gather evidence to determine if the assumption is warranted. A systematic review was conducted to this end (chapter 2) with two main findings.

Firstly, there was interest in questions pertaining to how and why an intervention work and secondly, these questions were often addressed using qualitative methods. The use of qualitative methods precludes a formal testing of the hypotheses about how and why an intervention works. The absence of formal testing of hypotheses also meant that there is a lack of evidence to back up any claims of the *hows* and *whys*.

Given the lack of use of causal inference methods and the clear interest in questions of *how* and *why*, it is likely that methodological barriers exist in the application of causal inference methods. A review of existing methods for estimating mediated causal effects, a common way in which questions of *how* and *why* manifest themselves, indicate that existing methods cater to a small subset of scenarios commonly encountered by researchers. Most notably, two of such gaps, the lack of estimation methods for causal odds ratio (**OR**) and for multiple mediators in sequence, were singled out for further development in this thesis.

The first step to addressing these gaps was to adopt a framework for causal inference to work in. This was addressed in chapter 3. Briefly, the potential outcome (**PO**) framework was adopted because the experimental design of the randomised controlled trial (**RCT**), long used to address questions of causality can be conceptualised as a comparison of **PO**. This provided a firm foundation to define the causal effects of interests and for this thesis: the mediated causal **OR** and the sequentially mediated causal effects. The framework was also used by other researchers as to consider questions of causality. Notably, Pearl (2001) used it to develop the mediation formula where mediated causal effects can be defined using expectations and probabilities. This was notable because it separates out the definition of the mediated causal effect from its methods of estimation. With the mediation formula, one can then use a model most suitable for the problem at hand to estimate the expectations and probabilities that exist in the definition of the mediated causal effect. Once these had been estimated, they can then be put together using the formula to obtain an estimate of the mediated causal effect. The way in which this was done was discussed in chapter 4. In chapter 5, methods and results of a validation procedure used to confirm that the causal effects were estimated as defined was discussed.

With the causal effects estimated, one last question remains. This relates to the assumption of no confounding between the mediator and the outcome. Had there been confounding between the mediator and the outcome, i.e. there is a common cause for both the mediator and the outcome,

then the effects of the mediator on the outcome may not be due solely to the mediator. This assumption is required because it is not possible to distinguish between the presence or absence of confounding since we do not know all possible common causes. Chapter 6 considers the scenarios where there are different amounts of confounding. A set of tests were developed to determine the extent to which the outcome would change had there been different amounts of confounding. Additionally, data from the Carers' Assessment, Skills and Information Sharing (**CASIS**) trial was used to demonstrate the use of the methods developed.

7.2 Strengths and limitations of methods developed.

The previous section briefly provided an overview of the thesis. This section discusses some of its strengths and weaknesses. There are three strengths and a weakness of this thesis that I will discuss.

Firstly, the thesis includes a chapter on validation of its results (chapter 5). This validation goes beyond theoretical correctness of the methods presented in chapter 4 to show empirically that the results are congruent with the estimates that would have been obtained using the well-established model-based approaches. While a valid criticism of such an approach towards validation is that it was done by using another estimator (the model-based estimator) to check the correctness of the novel simulation-based estimator, this was done because the manner in which the mediated causal effects were estimated using the simulation-based estimator builds upon the same building blocks as the model-based estimators. Furthermore, this validation is building on the model-based estimator, an estimator that had been in use for the better part of the past forty years. It also meant that whatever we know about the model-based estimator, can similarly be used when considering the novel estimator. Examples of these include the appropriate confidence interval to use. The Wald's confidence interval applied for the model-based estimator is known to be biased for the indirect effects. The same is true for this novel estimator and building on this prior knowledge, more appropriate confidence intervals constructed using the simulated causal effects were used.

Secondly, this thesis uniquely allows one to conduct sensitivity analysis for causal **OR**. This was made possible because *Mplus* provided the functionality to estimate logistic regression

models which includes latent variables. While it is regrettable that freely available software was not available to conduct such estimations, the ability to conduct sensitivity analysis for causal **OR** represents an improvement over existing methods. Other packages commonly used in the estimation of latent variable models such as *laavan* and *OpenMx* do not currently have the capability but through personal communication with the author of *lavvaan*, this is work in progress.

Thirdly, this thesis made use of a quasi-Bayesian simulation method originally developed by King et al. (2000) to conduct the causal effect estimation. This thesis is not the first to utilise such simulations but the development of the causal **OR** serves as a proof of concept that the simulation method is flexible and can be used to conduct estimations of causal effects for other parametric models. This is the case as long as we have a valid mechanism to use the estimated regression model as a data generating mechanism which covers models in the generalised linear model (**GLM**) family. Furthermore, should it be non-trivial to use the estimated model as a data generating mechanism, one can turn to bootstrap as a general way of conducting the data generation. This had been implemented in the algorithm and while it had not been scrutinised at the same level as the parametric based solution, the theoretical concept of using bootstrap as a data generating mechanism is a valid one as suggested both by Imai et al. (2010) and King et al. (2000).

Moving on to the weakness of this thesis, the estimation of causal effects assumes that there are no interaction effects between any of the variables. This limitation is due to a lack of time rather than a limitation of the method. Imai et al. (2010) in the *mediation* package allowed for interaction effects which also indicates that it is possible to do so. The difficulty in incorporating interaction effects comes mainly from a modification of the relationship between the total effects (**TE**), direct effect (**DE**) and indirect effect (**IE**). Should interaction effects be incorporated into the novel estimator, one would also need to consider how the relationship between the effects would change in the case of the causal **OR**. This could be addressed in a next revision of the simulation-based estimator.

7.3 Significant contributions

This section highlights significant contributions to the estimation of causal effects in the course of developing the estimators. These contributions differ from what was discussed previously and focuses on uses of the applicability of the methods developed in this thesis in other areas. There are three contributions that are notable.

7.3.1 Review of use of analytic methodology in current literature

The first significant contribution is in the methods used for the review of the use of analytic methods in current literature. Most systematic reviews concern with the assessment of the evidence supporting an observation and the methods developed for reviewing literature had mostly focused on this use case. The novel contribution in this thesis is to use and adapt the methods developed for systematic reviews of efficacies or effectiveness of therapeutics for the purposes of reviewing the current use of analytic methods for a class of problem in a specific domain. This builds upon the same principles used by systematic reviews and shares procedures used to ensure rigour of the review. This includes having a second reviewer and ensuring reproducibility of the review. Additionally, methods were also developed to study how a concept is phrased in current literature to ensure that important studies are not left out due to a lack of awareness of different ways in which a concept is expressed. Put together, the contributions provides methods to review the use of methods for analysis and a set of procedures that can be used to identify the varied ways in which a concept can be expressed.

7.3.2 Extension of the mediation formula

The second significant contribution is the development of an extension of the mediation formula. The mediation formula, conceptualised by Pearl (1995), is a formula that enabled one to obtain an expression representing a mediated causal effect. The mediation formula is notably a non-parametric formula that allows one to obtain an expression for a mediated causal effect of interest using only the definition of the causal effect as its input. The mediation formula expressed

the mediated causal effects using expectations and probability but does not specify how the expectations or probabilities are to be estimated. With the causal effects expression, one then need to estimate the quantities represented by the expression to obtain a causal effect estimate. For both the problems tackled in this thesis, namely, the estimation of sequentially mediated causal effects and the estimation of causal **OR**, I needed to use the mediation formula to obtain an expression for the mediated causal effect. In the former, the application of the mediation formula is straightforward but in the latter, the mediation formula cannot be applied directly because the mediation formula only catered to the scenario of a single mediator. In order to obtain an expression for causal effects for sequentially mediated causal effects with more than a single mediator, the mediation formula needs to be extended.

The mediation formula exists in various forms depending on what needs to be estimated. If the outcome is binary, there is a binary version of the mediation formula. However, regardless of what form the mediation formula takes, it rests upon the same underlying concept. The different versions of the formulae are simplifications of the original formulae and this simplification is different depending on whether outcome is estimated as a probability or as a measure of the outcome itself.

Consider a mediated causal effect estimated as a difference between two continuous **PO**. The expression of each of the **PO** is an application of the mediation formula and in the case of a single mediator, there will be a nested mediator **PO** within the expression of the outcome model. In order to estimate the expectation of the outcome model, one would need to consider what to do with the nested mediator. This is where the 'Tower's Law' or more formally known as the 'law of iterated expectations', can be applied to obtain an estimate of the expectation of the **PO**. The law works by computing a weighted sum of all possible values of the mediator nested within the outcome expression. In practice, this means to first obtain all possible values of the mediator, substitute them into the expression for the outcome. Each of the outcome expression containing a different value of the mediator is then weighted according to the probability by which they occur and then all the outcome expressions are then summed up. This procedure allows one to obtain the expression for estimating a single **PO**. A complication arises with more than a single mediator. Consider the scenario for two sequentially ordered mediators. The first mediator contains no nested terms while the second mediator has the first mediator nested within it and the outcome has both the first and second mediator nested within. In order to

obtain the expression to estimate a given **PO**, the expression has to be built up in a specific order. For the two scenario case, the expression for the first mediator has to be derived because it is the only dependent variable with no nested terms. Once the expression for the first mediator is obtained, we can then use this expression to obtain the expression for the second mediator using the law of iterated expectation. With the expression for the first and second mediator, we can then use them to obtain the expression for the outcome. The complication arises because the weighted sum of all possible values of the mediator in the single mediator scenario will now mean all possible *combinations of values* of all mediators. This has to be generated together with the corresponding probability of the combination occurring before the weighted sum could be obtained. The scenario for more than two mediators get much more complex with both the order in which the expressions for the mediators and the combinations for the nested terms in each of the dependent variables needing to be derived in the manner stated before. The contribution of this thesis is developing a general algorithm to obtain the expression for any number of sequentially ordered mediators. This done by recursively applying the law of iterated expectation until expressions for all the mediators and outcomes can be obtained. This also highlights a general way by which the mediation formula can be extended through a recursive application of the law of iterated expectation. This points towards a way by which mediation estimators for more complex scenarios might be obtained.

This extension of the mediation formula also gave rise to an insight. Since the law states that the expectation of the **PO** is a weighted sum of all possible values of the nested **PO**, in the case of a binary nested **PO**, the weighted sum of the **PO** is the expectation if we made infinite draws of the nested **PO** and summing up the weighted sum of the binary outcomes. In the context of a simulation, this also meant that if I were to conduct the simulation infinite number of times, each time realising the nested **PO**, I would obtain the weighted sum of the nested **PO**. The weighted sum of the **PO** fully took into account the fundamental error of the nested **PO**. Given that we can fully account for the fundamental error of the nested **PO**, in the context of a simulation, we can therefore not simulate and realise the binary nested **PO** and instead use the weighted sum. This is a deviation from prior work by Imai et al. (2010) in the 'mediation' R package and allows for the simulation to be completed more quickly.

7.3.3 Adaptation of quasi-Bayesian simulation method

The third significant contribution is the adaptation of the quasi-Bayesian simulation method for estimation of causal effects. As suggested previously, this adaptation is a proof of concept for how the simulation method can be used to simulate **PO** as long as we can use the estimated models as data-generating mechanisms. This simulation method can also be more widely used in comparisons of **PO** in other areas such as policy analysis. Questions such as 'given the known relationship between tobacco taxation and health outcomes, what is the reduction of harms associated with a given increase in taxation?'. Such questions can be framed as a comparison of **PO**s, much like how the estimation of causal effects requires the comparison of two **PO**. This implies that the method used to estimate causal effects framed as a comparison of **PO**s can similarly be used to evaluate such 'what-ifs' questions. The work done in this thesis furthered methods for use in such comparisons by adapting it for the logistic regression case. What is important is the general applicability of such an adaptation where a data generating mechanism is identified and used to simulate the **PO**. The mechanism used in this thesis taps on the paramteric nature of the models used but the data generating mechanism is not restricted to this mechanism only. The mechanism used in this thesis considers the roles of the two forms of uncertainty, the estimation and the fundamental uncertainty to simulate possible values of the **PO** under different conditions. Other mechanisms could be developed to address different sources of uncertainty thus allowing the simulated **PO** to reflect the uncertainties that is relevant to the question on hand.

Another use of the quasi-Bayesian simulation as demonstrated in this thesis is to evaluate integrals that have no known closed form solutions. This meant that the evaluation of the **PO** is now no longer constrained or dependent upon having a closed form solution of the expression derived from the application of the mediation formula. The use of the quasi-Bayesian simulation in evaluating integrals is an added advantage over the use of more traditional ways of conducting computer assisted integration by not requiring one to write out the expression of the integral one wishes to estimate. This advantage thus lowers the bar for researchers to apply methods that evaluates integrals without having to state the precise form of the integral because this evaluation of the integral is conceptualised as a mixing of two distributions.

7.4 Future developments and concluding remarks

This thesis focused on two main problems. The estimation of mediated causal **OR** and the sequentially mediated causal effects. While these two problems cover a wide range of scenarios often encountered in research, there is room for further development. One area is to allow the estimation to take into account interaction effects. This requires a redefinition of the relationships between the **TE** and the direct and indirect effects. The sensitivity analysis for such an estimation is also required.

Another area of further development is the estimation of a mediated complier average causal effect (**CACE**). **CACE** is an estimator that is used to estimate causal effects for the treatment received rather than the treatment offered as was done for this thesis. This involves a consideration of how best to conceptualise the **CACE** for it to fit into the existing estimation concepts. The utility of this is similar to the need for **CACE**, i.e. to understand the causal relationships between treatment received and the outcomes of interest.

Other areas that can be further developed includes applying the estimation methods for other **GLM** models. As was demonstrated in this thesis in the development of the causal **OR**, application of the estimation methods while possible, likely involves careful consideration of the appropriate data generating mechanism as well as ways by which the simulated **PO** is to be combined.

Lastly, the work conducted in this thesis aimed to assist researchers in the application of causal inference methods and to that end, the methods developed have bridged some gaps in methodology. More work will need to be done to address the different scenarios that are encountered by researchers.

Appendices

Appendix A

Keyword study

The keyword study starts off with the phrase 'process evaluation'. The root of the words, 'process' and 'evaluation' combined with keyword truncation ('process*' and 'evaluat*') (U.S. National Library of Medicine, 2018) were used as search terms using the search engine for Pubmed hosted at the National Center for Biotechnology Information (**NCBI**). This was done programatically through *R* (R Core Team, 2018) package, *RISmed* (Kovalchik, 2017).

The results from the search engine not only returned the studies that matched the search terms, it also conducted a 'query translation' that enumerated all the variations of the words 'process' and 'evaluation' (Appendix A.3 and Appendix A.2). The variations of the two words were permutated and the resulting list of phrases was searched programatically.

A total of 57462 possible phrase permutations were searched of which only 16 of the permutations returned any search results. The number of results for each permutation is shown in Appendix A.1. The top 4 permutations are plural forms of 'process evaluation' while the rest of the phrases with hits were not referring to **PE** in the same context as the current review. Therefore, only the first four permutations were used for subsequent searches. The abstracts of permutations with more than one result were reviewed to identify any alternative expression of the concept of **PE**. This did not yield any additional relevant keywords. The relevant identified variations of the keywords were then formulated as search terms for use in the subsequent searches. This thus conclude the keyword study.

Table A.1: Permutation of variations and their respective hits. Permutations with no hits are not shown.

Word variation	Hits
process evaluation	1662
evaluation process	1621
evaluation processes	288
process evaluations	159
evaluative processes	82
evaluative process	61
evaluative processing	51
evaluating process	24
process evaluated	22
processing evaluation	20
evaluating processes	17
evaluated processes	13
process evaluating	11
evaluation processing	9
process evaluator	8
processing evaluations	3
processual evaluation	1

Table A.2: Variations of ‘process’

process	process’	process”
process’in	process’resembling	process’s
process1	process7	processa
processabilities	processability	processable
processable’	processablity	processacceptability
processadas	processado	processadora
processadoras	processados	processalso
processamento	processand	processappeared
processare	processbased	processbearing
processblack	processcd	processclassical
processcompatible	processconsist	processd
processdb	processdiagnostics	processdiffraction
processe	processeable	processed
processed’	processedby	processeddata
processedfor	processeffect	processen
processequals	processer	processers
processes	processes’	processes1
processes3	processes3,4	processes4
processesacting	processesand	processesd
processesed	processeses	processesin
processesing	processesinvolved	processesnecessary
processesof	processess	processeswas
processesx	processeta	processew
processfrom	processg	processgenelists

Table A.2: Variations of ‘process’ (cont’d)

processguide	processhas	processi
processiable	processibility	processible
processiblity	processid	processidae
processig	processiing	processin
processincidenceprognosispredictive	processincidenceprognostic	processindicators
processing	processing’	processing”
processing’s	processing1	processing439
processingand	processingby	processingdependent
processingenzymes	processingg	processingocellar
processingperspectives	processings	processings’
processingt	processingtrade	processins
processinvolving	procession	processiona
processionae	processional	processionalis
processionals	processionals’	processionary
processione	processionea	processions
processis	processisng	processitivity
processive	processive’	processively
processivelyalong	processiveness	processivities
processivity	processlike	processmediated
processment	processness	processo
processoccurs	processodi	processof
processome	processomes	processomics
processone	processor	processor’
processor’s	processore	processori

Table A.2: Variations of ‘process’ (cont’d)

processors	processors’	processors’s
processos	processosome	processou
processov	processparameters	processphantasies
processpsycinfo	processretained	processs
processsed	processses	processsgp
processsignificant	processsing	processsus
processthe	processthis	processtivity
processtrade	processtreatment	processu
processual	processualised	processualism
processualist	processuality	processually
processural	processus	processuses
processuusing	processvia	processwas
processwater	processwere	processwith

Table A.3: Variations of ‘evaluation’

evaluat	evaluatability	evaluatable
evaluataion	evaluatation	evaluatble
evaluatc	evaluatd	evaluate
evaluate’	evaluatea	evaluateclinical
evaluated	evaluated’	evaluated4
evaluatedafter	evaluatedand	evaluatedbased
evaluatedble	evaluatedby	evaluateded
evaluatedfor	evaluatedfurther	evaluatedi
evaluatedin	evaluatedpatients	evaluatedsignificantly
evaluatedsix	evaluatedthe	evaluatedthrough
evaluatedusing	evaluatedwith	evaluatedwiththe
evaluatee	evaluatees	evaluatees’
evaluategenetic	evaluatein	evaluateing
evaluatematernal	evaluatenatural	evaluatepharma’s
evaluatepost	evaluater	evaluatereplacement
evaluaters	evaluates	evaluatesthe
evaluatestheir	evaluatet	evaluatethe
evaluatetm	evaluateur	evaluateurs
evaluatezseverity	evaluati	evaluatie
evaluatieperiode	evaluaties	evaluatif
evaluatiing	evaluatiion	evaluatijon
evaluatin	evaluating	evaluating’
evaluatinginflammation	evaluatingoutcomes	evaluatingprimary
evaluatingrhizophagus	evaluatingthe	evaluatins

Table A.3: Variations of ‘evaluation’ (cont’d)

evaluatio	evaluatioin	evaluatiom
evaluation	evaluation’	evaluation”
evaluation’s	evaluation1	evaluational
evaluationand	evaluationappropriate	evaluationary
evaluationdagger	evaluatione	evaluationed
evaluationelective	evaluationen	evaluationfor
evaluationg	evaluationicuintensive	evaluationii
evaluationiii	evaluationin	evaluationism
evaluationism’	evaluationl	evaluationof
evaluationproblem	evaluations	evaluations’
evaluationsbogen	evaluationsdesigns	evaluationsergebnisse
evaluationsergebnissen	evaluationsfragebogen	evaluationsincluded
evaluationsinstrumente	evaluationsinterviews	evaluationskonzeptes
evaluationsmedium	evaluationsmethoden	evaluationsnoten
evaluationspraxis	evaluationsprogrammen	evaluationsprojekt
evaluationsprozess	evaluationsrevealed	evaluationss
evaluationsscore	evaluationssoftware	evaluationsstudie
evaluationsystems	evaluationszirkel	evaluationt
evaluationtions	evaluationtool	evaluationwas
evaluatioon	evaluatior	evaluative
evaluative’	evaluativejudgments	evaluatively
evaluativeness	evaluatives	evaluativism
evaluativism’	evaluativist	evaluativistic
evaluativo	evaluativofueron	evaluativos

Table A.3: Variations of ‘evaluation’ (cont’d)

evaluatlion	evaluatng	evaluaton
evaluator	evaluator’	evaluator’s
evaluators	evaluators’	evaluatortrade
evaluatory	evaluats	evaluatuon
evaluatyed		

Appendix B

Search terms and specifications

B.1 Search engine: Wiley

B.1.1 Database: CENTRAL

```
+-----------------------------------------------------------------+
# Subject filters
+-----------------------------------------------------------------+
## INCLUDED
"process* evaluat*":ti,ab
## EXCLUDED
"protocol":ti
## LIMITS
Publication Year from 2007 to 2016, in Trials
+-----------------------------------------------------------------+
# Combined
+-----------------------------------------------------------------+
"process* evaluat*":ti,ab NOT "protocol":ti
Publication Year from 2012 to 2016, in Trials
+-----------------------------------------------------------------+
# Results
+-----------------------------------------------------------------+
Retrieved on 2017-02-27
328 hits
+-----------------------------------------------------------------+
```

B.2 Search engine: EBSCOhost

B.2.1 Database: CINAHL

```
+-------------------------------------------------------------------+
# RCT filters
+-------------------------------------------------------------------+
## INCLUDED
(PT "clinical trial")
(PT "clinical trial, phase i")
(PT "clinical trial, phase ii")
(PT "clinical trial, phase iii")
(PT "clinical trial, phase iv")
(PT "controlled clinical trial")
(PT "randomized controlled trial")
(PT "pragmatic clinical trial")
(MH "clinical trials as topic+")
((TI "clinical trial*") OR (AB "clinical trial*"))
((TI "trial*") OR (AB "trial*"))
(MW "DT")

(MH "random allocation+")
(TI (("allocat*" OR "assign*") N2 "random*") OR AB (("allocat*" OR "assign*")
N2 "random*"))
((TI "randomi?ed") OR (AB "randomi?ed"))
((TI "randomly") OR (AB "randomly"))
(TI "rct") OR (AB "rct")

(MH "placebos+")
((TI "placebo*") OR (AB "placebo*"))

(MH "prospective studies+")

(PT "multicenter study")
(MH "multicenter studies as topic+")
((TI "groups") OR (AB "groups"))

(MH "cross-over studies+")
(TI ("crossover*" OR "cross over*" OR "cross-over*") OR AB ("crossover*" OR
"cross over*" OR "cross-over*"))

(MH "double blind method+")
```

```
(MH "single blind method+")
((TI (("singl*" OR "doubl*" OR "treb*" OR "tripl*") N1 ("blind*" OR "mask*")))
OR "(AB (("singl*" OR "doubl*" OR "treb*" OR "tripl*") N1 ("blind*" OR
"mask*"))))
## EXCLUDED
((MH "animals+") NOT (MH "humans+"))
(TI "case report*")
(PT "case reports")
(PT "letter")
(PT "historical article")
(MH "correspondence as topic")
+------------------------------------------------------------------+
# Subject filters
+------------------------------------------------------------------+
## INCLUDED
((TI "process* evaluat*") OR (AB "process* evaluat*"))
## EXCLUDED
(TI "protocol")
## LIMITS
(LA "English")

((PT "journal article") OR (PT "article"))

(DT "2012*" OR DT "2013*" OR DT "2014*" OR DT "2015*" OR DT "2016*")
(EM "2011*" OR EM "2012*" OR EM "2013*" OR EM "2014*" OR EM "2015*" OR EM
"2016*")

* Dates [CCYYMMDD]
DT - Date of publication
EM - Date created
+------------------------------------------------------------------+
# Combined
+------------------------------------------------------------------+
((((PT "clinical trial") OR (PT "clinical trial, phase i") OR (PT "clinical
trial, phase ii") OR (PT "clinical trial, phase iii") OR (PT "clinical trial,
phase iv") OR (PT "controlled clinical trial") OR (PT "randomized controlled
trial") OR (PT "pragmatic clinical trial") OR (MH "clinical trials as topic+")
OR ((TI "clinical trial*") OR (AB "clinical trial*")) OR ((TI "trial*") OR (AB
"trial*")) OR (MW "DT") OR (MH "random allocation+") OR (TI (("allocat*" OR
"assign*") N2 "random*") OR AB (("allocat*" OR "assign*") N2 "random*")) OR
((TI "randomi?ed") OR (AB "randomi?ed")) OR ((TI "randomly") OR (AB
"randomly")) OR (TI "rct") OR (AB "rct") OR (MH "placebos+") OR ((TI
"placebo*") OR (AB "placebo*")) OR (MH "prospective studies+") OR (PT
"multicenter study") OR (MH "multicenter studies as topic+") OR ((TI "groups")
OR (AB "groups")) OR (MH "cross-over studies+") OR (TI ("crossover*" OR "cross
```

```
over*" OR "cross-over*") OR AB ("crossover*" OR "cross over*" OR
"cross-over*")) OR (MH "double blind method+") OR (MH "single blind method+")
OR ((TI (("singl*" OR "doubl*" OR "treb*" OR "tripl*") N1 ("blind*" OR
"mask*")))" OR "(AB (("singl*" OR "doubl*" OR "treb*" OR "tripl*") N1 ("blind*"
OR "mask*"))))) NOT (((MH "animals+") NOT (MH "humans+")) OR (TI "case
report*") OR (PT "case reports") OR (PT "letter") OR (PT "historical article")
OR (MH "correspondence as topic") OR (PT "meeting abstracts") OR (PT
"abstracts"))) AND ((((TI "process* evaluat*") OR (AB "process* evaluat*")) NOT
(TI "protocol")) AND ((LA "English") AND ((PT "journal article") OR (PT
"article")) AND (DT "2012*" OR DT "2013*" OR DT "2014*" OR DT "2015*" OR DT
"2016*") AND (EM "2011*" OR EM "2012*" OR EM "2013*" OR EM "2014*" OR EM
"2015*" OR EM "2016*"))))
+-----------------------------------------------------------------------+
# Results
+-----------------------------------------------------------------------+
Retrieved on 2017-02-27
103 hits
+-----------------------------------------------------------------------+
```

B.2.2 Database: MEDLINE

```
+-----------------------------------------------------------------------+
# RCT filters
+-----------------------------------------------------------------------+
## INCLUDED
(PT "clinical trial")
(PT "clinical trial, phase i")
(PT "clinical trial, phase ii")
(PT "clinical trial, phase iii")
(PT "clinical trial, phase iv")
(PT "controlled clinical trial")
(PT "randomized controlled trial")
(PT "pragmatic clinical trial")
(MH "clinical trials as topic+")
((TI "clinical trial*") OR (AB "clinical trial*"))
((TI "trial*") OR (AB "trial*"))
(MW "DT")

(MH "random allocation+")
(TI (("allocat*" OR "assign*") N2 "random*") OR AB (("allocat*" OR "assign*")
N2 "random*"))
((TI "randomi?ed") OR (AB "randomi?ed"))
((TI "randomly") OR (AB "randomly"))
```

```
(TI "rct") OR (AB "rct")

(MH "placebos+")
((TI "placebo*") OR (AB "placebo*"))

(MH "prospective studies+")

(PT "multicenter study")
(MH "multicenter studies as topic+")
((TI "groups") OR (AB "groups"))

(MH "cross-over studies+")
(TI ("crossover*" OR "cross over*" OR "cross-over*") OR AB ("crossover*" OR
"cross over*" OR "cross-over*"))

(MH "double blind method+")
(MH "single blind method+")
((TI (("singl*" OR "doubl*" OR "treb*" OR "tripl*") N1 ("blind*" OR "mask*")))"
OR "(AB (("singl*" OR "doubl*" OR "treb*" OR "tripl*") N1 ("blind*" OR
"mask*"))))
## EXCLUDED
((MH "animals+") NOT (MH "humans+"))
(TI "case report*")
(PT "case reports")
(PT "letter")
(PT "historical article")
(MH "correspondence as topic")
+-----------------------------------------------------------------+
# Subject filters
+-----------------------------------------------------------------+
## INCLUDED
((TI "process* evaluat*") OR (AB "process* evaluat*"))
## EXCLUDED
(TI "protocol")
## LIMITS
(LA "English")

((PT "journal article") OR (PT "article"))

(DT "2012*" OR DT "2013*" OR DT "2014*" OR DT "2015*" OR DT "2016*")
(EM "2011*" OR EM "2012*" OR EM "2013*" OR EM "2014*" OR EM "2015*" OR EM
"2016*")

* Dates [CCYYMMDD]
DT - Date of publication
```

```
EM - Date created
+-----------------------------------------------------------------------+
# Combined
+-----------------------------------------------------------------------+
((((PT "clinical trial") OR (PT "clinical trial, phase i") OR (PT "clinical
trial, phase ii") OR (PT "clinical trial, phase iii") OR (PT "clinical trial,
phase iv") OR (PT "controlled clinical trial") OR (PT "randomized controlled
trial") OR (PT "pragmatic clinical trial") OR (MH "clinical trials as topic+")
OR ((TI "clinical trial*") OR (AB "clinical trial*")) OR ((TI "trial*") OR (AB
"trial*")) OR (MW "DT") OR (MH "random allocation+") OR (TI (("allocat*" OR
"assign*") N2 "random*") OR AB (("allocat*" OR "assign*") N2 "random*")) OR
((TI "randomi?ed") OR (AB "randomi?ed")) OR ((TI "randomly") OR (AB
"randomly")) OR (TI "rct") OR (AB "rct") OR (MH "placebos+") OR ((TI
"placebo*") OR (AB "placebo*")) OR (MH "prospective studies+") OR (PT
"multicenter study") OR (MH "multicenter studies as topic+") OR ((TI "groups")
OR (AB "groups")) OR (MH "cross-over studies+") OR (TI ("crossover*" OR "cross
over*" OR "cross-over*") OR AB ("crossover*" OR "cross over*" OR
"cross-over*")) OR (MH "double blind method+") OR (MH "single blind method+")
OR ((TI (("singl*" OR "doubl*" OR "treb*" OR "tripl*") N1 ("blind*" OR
"mask*")))" OR "(AB (("singl*" OR "doubl*" OR "treb*" OR "tripl*") N1 ("blind*"
OR "mask*"))))) NOT (((MH "animals+") NOT (MH "humans+")) OR (TI "case
report*") OR (PT "case reports") OR (PT "letter") OR (PT "historical article")
OR (MH "correspondence as topic") OR (PT "meeting abstracts") OR (PT
"abstracts"))) AND ((((TI "process* evaluat*") OR (AB "process* evaluat*")) NOT
(TI "protocol")) AND ((LA "English") AND ((PT "journal article") OR (PT
"article")) AND (DT "2012*" OR DT "2013*" OR DT "2014*" OR DT "2015*" OR DT
"2016*") AND (EM "2011*" OR EM "2012*" OR EM "2013*" OR EM "2014*" OR EM
"2015*" OR EM "2016*"))))
+-----------------------------------------------------------------------+
# Results
+-----------------------------------------------------------------------+
Retrieved on 2017-02-27
564 hits
+-----------------------------------------------------------------------+
```

B.3 Search engine: Embase.com

B.3.1 Database: Embase

```
+-----------------------------------------------------------------------+
# RCT filters
```

```
+---------------------------------------------------------------------+
## INCLUDED
('clinical trial'/exp)
('clinical trial (topic)'/exp)
('clinical trial*':ti,ab)
(('trial*':ti,ab))
('drug therapy':lnk)

('randomization'/exp)
((('allocat*' OR 'assign*') NEAR/2 'random*'):ti,ab)
('randomi*ed':ti,ab)
('randomly':ti,ab)
(('rct'):ti,ab)

('crossover procedure'/exp)
(('crossover*' OR 'cross over*' OR 'cross-over*'):ti,ab)

('placebos'/exp)
('placebo*':ti,ab)

('prospective study'/exp)

('groups':ti,ab)

('double blind procedure'/exp)
('single blind procedure'/exp)
((('singl*' OR 'doubl*' OR 'tripl*' OR 'trebl*') NEAR/1 ('blind*' OR
'mask*')):ti,ab)
## EXCLUDED
('animals'/exp) NOT ('humans'/exp)
('case report*':ti)
('case study'/exp)
('letter'/exp)
('historical research'/exp)
('history of medicine'/exp)
('abstract report'/exp)
+---------------------------------------------------------------------+
# Subject filters
+---------------------------------------------------------------------+
## INCLUDED
(('process* evaluat*'):ti,ab)
## EXCLUDED
(('protocol'):ti)
## LIMITS
([article]/lim OR [article in press]/lim OR [review]/lim)
```

```
NOT ([afrikaans]/lim OR [albanian]/lim OR [arabic]/lim OR [armenian]/lim OR
[azerbaijani]/lim OR [basque]/lim OR [belarusian]/lim OR [bengali]/lim OR
[bosnian]/lim OR [bulgarian]/lim OR [burmese]/lim OR [catalan]/lim OR
[chinese]/lim OR [croatian]/lim OR [czech]/lim OR [danish]/lim OR [dutch]/lim
OR [esperanto]/lim OR [estonian]/lim OR [finnish]/lim OR [french]/lim OR
[georgian]/lim OR [german]/lim OR [greek]/lim OR [hebrew]/lim OR [hindi]/lim OR
[hungarian]/lim OR [icelandic]/lim OR [indonesian]/lim OR [irish gaelic]/lim OR
[italian]/lim OR [japanese]/lim OR [korean]/lim OR [latvian]/lim OR
[lithuanian]/lim OR [macedonian]/lim OR [malay]/lim OR [mongolian]/lim OR
[norwegian]/lim OR [persian]/lim OR [polish]/lim OR [polyglot]/lim OR
[portuguese]/lim OR [pushto]/lim OR [romanian]/lim OR [russian]/lim OR
[scottish gaelic]/lim OR [serbian]/lim OR [sinhalese]/lim OR [slovak]/lim OR
[slovenian]/lim OR [spanish]/lim OR [swedish]/lim OR [tagalog]/lim OR
[thai]/lim OR [turkish]/lim OR [ukrainian]/lim OR [urdu]/lim OR [uzbek]/lim OR
[vietnamese]/lim)
([2012-2016]/py)
([1-1-2011]/sd NOT [31-12-2016]/sd)
([embase]/lim)
+---------------------------------------------------------------------+
# Combined
+---------------------------------------------------------------------+
((((('clinical trial'/exp) OR ('clinical trial (topic)'/exp) OR ('clinical
trial*':ti,ab) OR (('trial*':ti,ab)) OR ('drug therapy':lnk) OR
('randomization'/exp) OR ((('allocat*' OR 'assign*') NEAR/2 'random*'):ti,ab)
OR ('randomi*ed':ti,ab) OR ('randomly':ti,ab) OR (('rct'):ti,ab) OR ('crossover
procedure'/exp) OR (('crossover*' OR 'cross over*' OR 'cross-over*'):ti,ab) OR
('placebos'/exp) OR ('placebo*':ti,ab) OR ('prospective study'/exp) OR
('groups':ti,ab) OR ('double blind procedure'/exp) OR ('single blind
procedure'/exp) OR ((('singl*' OR 'doubl*' OR 'tripl*' OR 'trebl*') NEAR/1
('blind*' OR 'mask*')):ti,ab)) NOT (('animals'/exp) NOT ('humans'/exp) OR
('case report*':ti) OR ('case study'/exp) OR ('letter'/exp) OR ('historical
research'/exp) OR ('history of medicine'/exp) OR ('abstract report'/exp))) AND
((((('process* evaluat*'):ti,ab) NOT (('protocol'):ti)) AND (([article]/lim OR
[article in press]/lim OR [review]/lim) NOT ([afrikaans]/lim OR [albanian]/lim
OR [arabic]/lim OR [armenian]/lim OR [azerbaijani]/lim OR [basque]/lim OR
[belarusian]/lim OR [bengali]/lim OR [bosnian]/lim OR [bulgarian]/lim OR
[burmese]/lim OR [catalan]/lim OR [chinese]/lim OR [croatian]/lim OR
[czech]/lim OR [danish]/lim OR [dutch]/lim OR [esperanto]/lim OR [estonian]/lim
OR [finnish]/lim OR [french]/lim OR [georgian]/lim OR [german]/lim OR
[greek]/lim OR [hebrew]/lim OR [hindi]/lim OR [hungarian]/lim OR
[icelandic]/lim OR [indonesian]/lim OR [irish gaelic]/lim OR [italian]/lim OR
[japanese]/lim OR [korean]/lim OR [latvian]/lim OR [lithuanian]/lim OR
[macedonian]/lim OR [malay]/lim OR [mongolian]/lim OR [norwegian]/lim OR
[persian]/lim OR [polish]/lim OR [polyglot]/lim OR [portuguese]/lim OR
[pushto]/lim OR [romanian]/lim OR [russian]/lim OR [scottish gaelic]/lim OR
```

```
[serbian]/lim OR [sinhalese]/lim OR [slovak]/lim OR [slovenian]/lim OR
[spanish]/lim OR [swedish]/lim OR [tagalog]/lim OR [thai]/lim OR [turkish]/lim
OR [ukrainian]/lim OR [urdu]/lim OR [uzbek]/lim OR [vietnamese]/lim) AND
([2012-2016]/py) AND ([1-1-2011]/sd NOT [31-12-2016]/sd) AND ([medline]/lim))))
+-----------------------------------------------------------------+
# Results
+-----------------------------------------------------------------+
Retrieved on 2017-02-27
227 hits
+-----------------------------------------------------------------+
```

B.3.2 Database: MEDLINE

```
+-----------------------------------------------------------------+
# RCT filters
+-----------------------------------------------------------------+
## INCLUDED
('clinical trial'/exp)
('clinical trial (topic)'/exp)
('clinical trial*':ti,ab)
(('trial*':ti,ab))
('drug therapy':lnk)

('randomization'/exp)
((('allocat*' OR 'assign*') NEAR/2 'random*'):ti,ab)
('randomi*ed':ti,ab)
('randomly':ti,ab)
(('rct'):ti,ab)

('crossover procedure'/exp)
(('crossover*' OR 'cross over*' OR 'cross-over*'):ti,ab)

('placebos'/exp)
('placebo*':ti,ab)

('prospective study'/exp)

('groups':ti,ab)

('double blind procedure'/exp)
('single blind procedure'/exp)
((('singl*' OR 'doubl*' OR 'tripl*' OR 'trebl*') NEAR/1 ('blind*' OR
'mask*')):ti,ab)
```

```
## EXCLUDED
('animals'/exp) NOT ('humans'/exp)
('case report*':ti)
('case study'/exp)
('letter'/exp)
('historical research'/exp)
('history of medicine'/exp)
('abstract report'/exp)
+---------------------------------------------------------------------+
# Subject filters
+---------------------------------------------------------------------+
## INCLUDED
(('process* evaluat*'):ti,ab)
## EXCLUDED
(('protocol'):ti)
## LIMITS
([article]/lim OR [article in press]/lim OR [review]/lim)
NOT ([afrikaans]/lim OR [albanian]/lim OR [arabic]/lim OR [armenian]/lim OR
[azerbaijani]/lim OR [basque]/lim OR [belarusian]/lim OR [bengali]/lim OR
[bosnian]/lim OR [bulgarian]/lim OR [burmese]/lim OR [catalan]/lim OR
[chinese]/lim OR [croatian]/lim OR [czech]/lim OR [danish]/lim OR [dutch]/lim
OR [esperanto]/lim OR [estonian]/lim OR [finnish]/lim OR [french]/lim OR
[georgian]/lim OR [german]/lim OR [greek]/lim OR [hebrew]/lim OR [hindi]/lim OR
[hungarian]/lim OR [icelandic]/lim OR [indonesian]/lim OR [irish gaelic]/lim OR
[italian]/lim OR [japanese]/lim OR [korean]/lim OR [latvian]/lim OR
[lithuanian]/lim OR [macedonian]/lim OR [malay]/lim OR [mongolian]/lim OR
[norwegian]/lim OR [persian]/lim OR [polish]/lim OR [polyglot]/lim OR
[portuguese]/lim OR [pushto]/lim OR [romanian]/lim OR [russian]/lim OR
[scottish gaelic]/lim OR [serbian]/lim OR [sinhalese]/lim OR [slovak]/lim OR
[slovenian]/lim OR [spanish]/lim OR [swedish]/lim OR [tagalog]/lim OR
[thai]/lim OR [turkish]/lim OR [ukrainian]/lim OR [urdu]/lim OR [uzbek]/lim OR
[vietnamese]/lim)
([2012-2016]/py)
([1-1-2011]/sd NOT [31-12-2016]/sd)
([medline]/lim)
+---------------------------------------------------------------------+
# Combined
+---------------------------------------------------------------------+
((((('clinical trial'/exp) OR ('clinical trial (topic)'/exp) OR ('clinical
trial*':ti,ab) OR (('trial*':ti,ab)) OR ('drug therapy':lnk) OR
('randomization'/exp) OR ((('allocat*' OR 'assign*') NEAR/2 'random*'):ti,ab) OR
('randomi*ed':ti,ab) OR ('randomly':ti,ab) OR (('rct'):ti,ab) OR ('crossover
procedure'/exp) OR (('crossover*' OR 'cross over*' OR 'cross-over*'):ti,ab) OR
('placebos'/exp) OR ('placebo*':ti,ab) OR ('prospective study'/exp) OR
('groups':ti,ab) OR ('double blind procedure'/exp) OR ('single blind
```

```
procedure'/exp) OR ((('singl*' OR 'doubl*' OR 'tripl*' OR 'trebl*') NEAR/1
('blind*' OR 'mask*')):ti,ab)) NOT (('animals'/exp) NOT ('humans'/exp) OR ('case
report*':ti) OR ('case study'/exp) OR ('letter'/exp) OR ('historical
research'/exp) OR ('history of medicine'/exp) OR ('abstract report'/exp))) AND
((((('process* evaluat*'):ti,ab) NOT (('protocol'):ti)) AND (([article]/lim OR
[article in press]/lim OR [review]/lim) NOT ([afrikaans]/lim OR [albanian]/lim
OR [arabic]/lim OR [armenian]/lim OR [azerbaijani]/lim OR [basque]/lim OR
[belarusian]/lim OR [bengali]/lim OR [bosnian]/lim OR [bulgarian]/lim OR
[burmese]/lim OR [catalan]/lim OR [chinese]/lim OR [croatian]/lim OR [czech]/lim
OR [danish]/lim OR [dutch]/lim OR [esperanto]/lim OR [estonian]/lim OR
[finnish]/lim OR [french]/lim OR [georgian]/lim OR [german]/lim OR [greek]/lim
OR [hebrew]/lim OR [hindi]/lim OR [hungarian]/lim OR [icelandic]/lim OR
[indonesian]/lim OR [irish gaelic]/lim OR [italian]/lim OR [japanese]/lim OR
[korean]/lim OR [latvian]/lim OR [lithuanian]/lim OR [macedonian]/lim OR
[malay]/lim OR [mongolian]/lim OR [norwegian]/lim OR [persian]/lim OR
[polish]/lim OR [polyglot]/lim OR [portuguese]/lim OR [pushto]/lim OR
[romanian]/lim OR [russian]/lim OR [scottish gaelic]/lim OR [serbian]/lim OR
[sinhalese]/lim OR [slovak]/lim OR [slovenian]/lim OR [spanish]/lim OR
[swedish]/lim OR [tagalog]/lim OR [thai]/lim OR [turkish]/lim OR [ukrainian]/lim
OR [urdu]/lim OR [uzbek]/lim OR [vietnamese]/lim) AND ([2012-2016]/py) AND
([1-1-2011]/sd NOT [31-12-2016]/sd) AND ([medline]/lim))))
+-----------------------------------------------------------------+
# Results
+-----------------------------------------------------------------+
Retrieved on 2017-02-27
335 hits
+-----------------------------------------------------------------+
```

B.4 Search engine: Ovid

B.4.1 Database: Embase

```
+-----------------------------------------------------------------+
# RCT filters
+-----------------------------------------------------------------+
## INCLUDED
((exp "clinical trial").sh.)
((exp "clinical trial (topic)").sh.)
(("clinical trial$").ti,ab.)
(("trial$").ti,ab.)
(("drug therapy").fs.)
```

```
((exp "randomization").sh.)
((("allocat$" OR "assign$") ADJ2 "random$").ti,ab.)
(("randomi#ed").ti,ab.)
(("randomly").ti,ab.)
(("rct").ti,ab.)

((exp "crossover procedure").sh.)
(("crossover$" OR "cross over$" OR "cross-over$").ti,ab.)

((exp "placebos").sh.)
(("placebo$").ti,ab.)

((exp "prospective study").sh.)

(("groups").ti,ab.)

((exp "double blind procedure").sh.)
((exp "single blind procedure").sh.)
((((("singl$" OR "doubl$" OR "treb$" OR "tripl$") ADJ1 ("blind$" OR
"mask$"))).ti,ab.)
## EXCLUDED
((exp "animals").sh. NOT (exp "humans").sh.)
(("case report$").ti.)
((exp "case study").sh.)
((exp "letter").sh.)
((exp "historical research").sh.)
((exp "history of medicine").sh.)
((exp "abstract report").sh.)
+---------------------------------------------------------------------+
# Subject filters
+---------------------------------------------------------------------+
## INCLUDED
(("process$" ADJ "evaluat$").ti,ab.)
## EXCLUDED
(("protocol").ti.)
## LIMITS
NOT (("afrikaans" OR "albanian" OR "arabic" OR "armenian" OR
"azerbaidzhani" OR "belorussian" OR "bosnian" OR "bulgarian" OR "catalan"
OR "chinese" OR "croatian" OR "czech" OR "danish" OR "dutch" OR "esperanto"
OR "estonian" OR "finnish" OR "french" OR "gallegan" OR "georgian" OR
"german" OR "greek" OR "hebrew" OR "hindi" OR "hungarian" OR "icelandic" OR
"indonesian" OR "irish gaelic" OR "italian" OR "japanese" OR "korean" OR
"latvian" OR "lithuanian" OR "macedonian" OR "malay" OR "norwegian" OR
"persian" OR "polish" OR "polyglot" OR "portuguese" OR "pushto" OR
```

```
"romanian" OR "russian" OR "scottish gaelic" OR "serbian" OR "sinhalese" OR
"slovak" OR "slovene" OR "spanish" OR "swedish" OR "thai" OR "turkish" OR
"ukrainian" OR "urdu" OR "uzbek" OR "vietnamese").lg.)

((2012$ OR 2013$ OR 2014$ OR 2015$ OR 2016$).dp.)
((2011$ OR 2012$ OR 2013$ OR 2014$ OR 2015$ OR 2016$).dc.)
+-----------------------------------------------------------------+
# Combined
+-----------------------------------------------------------------+
(((((exp "clinical trial").sh.) OR ((exp "clinical trial (topic)").sh.) OR
(("clinical trial$").ti,ab.) OR (("trial$").ti,ab.) OR (("drug
therapy").fs.) OR ((exp "randomization").sh.) OR ((("allocat$" OR
"assign$") ADJ2 "random$").ti,ab.) OR (("randomi#ed").ti,ab.) OR
(("randomly").ti,ab.) OR (("rct").ti,ab.) OR ((exp "crossover
procedure").sh.) OR (("crossover$" OR "cross over$" OR
"cross-over$").ti,ab.) OR ((exp "placebos").sh.) OR (("placebo$").ti,ab.)
OR ((exp "prospective study").sh.) OR (("groups").ti,ab.) OR ((exp "double
blind procedure").sh.) OR ((exp "single blind procedure").sh.) OR
((((("singl$" OR "doubl$" OR "treb$" OR "tripl$") ADJ1 ("blind$" OR
"mask$"))).ti,ab.)) NOT (((exp "animals").sh. NOT (exp "humans").sh.) OR
(("case report$").ti.) OR ((exp "case study").sh.) OR ((exp "letter").sh.)
OR ((exp "historical research").sh.) OR ((exp "history of medicine").sh.)
OR ((exp "abstract report").sh.))) AND ((((("process$" ADJ
"evaluat$").ti,ab.) NOT (("protocol").ti.)) AND (((2012$ OR 2013$ OR 2014$
OR 2015$ OR 2016$).dp.) AND ((2011$ OR 2012$ OR 2013$ OR 2014$ OR 2015$ OR
2016$).dc.) NOT (("afrikaans" OR "albanian" OR "arabic" OR "armenian" OR
"azerbaidzhani" OR "belorussian" OR "bosnian" OR "bulgarian" OR "catalan"
OR "chinese" OR "croatian" OR "czech" OR "danish" OR "dutch" OR "esperanto"
OR "estonian" OR "finnish" OR "french" OR "gallegan" OR "georgian" OR
"german" OR "greek" OR "hebrew" OR "hindi" OR "hungarian" OR "icelandic" OR
"indonesian" OR "irish gaelic" OR "italian" OR "japanese" OR "korean" OR
"latvian" OR "lithuanian" OR "macedonian" OR "malay" OR "norwegian" OR
"persian" OR "polish" OR "polyglot" OR "portuguese" OR "pushto" OR
"romanian" OR "russian" OR "scottish gaelic" OR "serbian" OR "sinhalese" OR
"slovak" OR "slovene" OR "spanish" OR "swedish" OR "thai" OR "turkish" OR
'a"ukrainian" OR "urdu" OR "uzbek" OR "vietnamese").lg.))))
+-----------------------------------------------------------------+
# Results
+-----------------------------------------------------------------+
Retrieved on 2017-02-27
169 hits
+-----------------------------------------------------------------+
```

B.4.2 Database: PsycINFO

```
+-------------------------------------------------------------------+
# RCT filters
+-------------------------------------------------------------------+
## INCLUDED
((exp "clinical trials").sh.)
(("clinical trial").md.)
(("clinical trial$").ti,ab.)
(("trial$").ti,ab.)
((exp "drug therapy").sh.)

((exp "random sampling").sh.)
((("allocat$" OR "assign$") ADJ2 "random$").ti,ab.)
(("randomi#ed").ti,ab.)
(("randomly").ti,ab.)
(("rct").ti,ab.)

((exp "placebo").sh.)
(("placebo$").ti,ab.)

((exp "prospective studies").sh.)
(("prospective study").md.)

(("groups").ti,ab.)

(("crossover$" OR "cross over$" OR "cross-over$").ti,ab.)

((((("singl$" OR "doubl$" OR "treb$" OR "tripl$") ADJ1 ("blind$" OR
"mask$"))).ti,ab.)

(("empirical study").md.)
## LIMITS
(("human").po.)
## EXCLUDED
(("case report$").ti.)
(("book").pt.)
(("encyclopedia").pt.)
+-------------------------------------------------------------------+
# Subject filters
+-------------------------------------------------------------------+
## INCLUDED
(("process$" ADJ "evaluat$").ti,ab.)
## EXCLUDED
```

```
(("protocol").ti.)
## LIMITS
NOT (("afrikaans" OR "albanian" OR "arabic" OR "bulgarian" OR "catalan" OR
"chinese" OR "croatian" OR "czech" OR "danish" OR "dutch" OR "estonian" OR
"farsi iranian" OR "finnish" OR "french" OR "georgian" OR "german" OR "greek"
OR "hebrew" OR "hindi" OR "hungarian" OR "italian" OR "japanese" OR "korean" OR
"lithuanian" OR "malaysian" OR "nonenglish" OR "norwegian" OR "polish" OR
"portuguese" OR "romanian" OR "russian" OR "serbian" OR "serbo croatian" OR
"slovak" OR "slovene" OR "spanish" OR "swedish" OR "turkish" OR "ukrainian" OR
"urdu").lg.)

((2012$ OR 2013$ OR 2014$ OR 2015$ OR 2016$).dp.)
((2011$ OR 2012$ OR 2013$ OR 2014$ OR 2015$ OR 2016$).up.)
+-----------------------------------------------------------------------+
# Combined
+-----------------------------------------------------------------------+
((((((exp "clinical trials").sh.) OR (("clinical trial").md.) OR (("clinical
trial$").ti,ab.) OR (("trial$").ti,ab.) OR ((exp "drug therapy").sh.) OR ((exp
"random sampling").sh.) OR ((("allocat$" OR "assign$") ADJ2 "random$").ti,ab.)
OR (("randomi#ed").ti,ab.) OR (("randomly").ti,ab.) OR (("rct").ti,ab.) OR
((exp "placebo").sh.) OR (("placebo$").ti,ab.) OR ((exp "prospective
studies").sh.) OR (("prospective study").md.) OR (("groups").ti,ab.) OR
(("crossover$" OR "cross over$" OR "cross-over$").ti,ab.) OR ((((singl$" OR
"doubl$" OR "treb$" OR "tripl$") ADJ1 ("blind$" OR "mask$"))).ti,ab.) OR
(("empirical study").md.)) AND (("human").po.)) NOT ((("case report$").ti.) OR
(("book").pt.) OR (("encyclopedia").pt.))) AND ((((("process$" ADJ
"evaluat$").ti,ab.) NOT (("protocol").ti.)) AND (((2012$ OR 2013$ OR 2014$ OR
2015$ OR 2016$).dp.) AND ((2011$ OR 2012$ OR 2013$ OR 2014$ OR 2015$ OR
2016$).up.) NOT (("afrikaans" OR "albanian" OR "arabic" OR "bulgarian" OR
"catalan" OR "chinese" OR "croatian" OR "czech" OR "danish" OR "dutch" OR
"estonian" OR "farsi iranian" OR "finnish" OR "french" OR "georgian" OR
"german" OR "greek" OR "hebrew" OR "hindi" OR "hungarian" OR "italian" OR
"japanese" OR "korean" OR "lithuanian" OR "malaysian" OR "nonenglish" OR
"norwegian" OR "polish" OR "portuguese" OR "romanian" OR "russian" OR "serbian"
OR "serbo croatian" OR "slovak" OR "slovene" OR "spanish" OR "swedish" OR
"turkish" OR "ukrainian" OR "urdu").lg.))))
+-----------------------------------------------------------------------+
# Results
+-----------------------------------------------------------------------+
Retrieved on 2017-02-27
404 hits
+-----------------------------------------------------------------------+
```

B.4.3 Database: MEDLINE

```
+---------------------------------------------------------------------+
# RCT filters
+---------------------------------------------------------------------+
## INCLUDED
((exp "clinical trial").pt.)
((exp "clinical trials as topic").sh.)
(("clinical trial$").ti,ab.)
(("trial$").ti,ab.)
(("drug therapy").fs.)

((exp "random allocation").sh.)
((("allocat$" OR "assign$") ADJ2 "random$").ti,ab.)
(("randomi#ed").ti,ab.)
(("randomly").ti,ab.)
(("rct").ti,ab.)

((exp "placebos").sh.)
(("placebo$").ti,ab.)

((exp "prospective studies").pt.)

((exp "multicenter study").pt.)
((exp "multicenter studies as topic").sh.)
(("groups").ti,ab.)

((exp "cross-over studies").sh.)
(("crossover$" OR "cross over$" OR "cross-over$").ti,ab.)

((exp "double blind method").sh.)
((exp "single blind method").sh.)
((((("singl$" OR "doubl$" OR "treb$" OR "tripl$") ADJ1 ("blind$" OR
"mask$"))).ti,ab.)
## EXCLUDED
((exp "animals").sh. NOT (exp "humans").sh.)
(("case report$").ti.)
(("case reports").pt.)
(("letter").sh.)
(("historical article").pt.)
(("correspondence as topic").sh.)
+---------------------------------------------------------------------+
# Subject filters
+---------------------------------------------------------------------+
```

```
## INCLUDED
(("process$" ADJ "evaluat$").ti,ab.)
## EXCLUDED
(("protocol").ti.)
## LIMITS
(("english" OR "eng" OR "und" OR "undetermined").lg.)

((2012$ OR 2013$ OR 2014$ OR 2015$ OR 2016$).dp.)
((2011$ OR 2012$ OR 2013$ OR 2014$ OR 2015$ OR 2016$).dc.)

* .dp. Date - Publication
* .dc. Date - Create
+-----------------------------------------------------------------+
# Combined
+-----------------------------------------------------------------+
(((((exp "clinical trial").pt.) OR ((exp "clinical trials as topic").sh.) OR
(("clinical trial$").ti,ab.) OR (("trial$").ti,ab.) OR (("drug therapy").fs.)
OR ((exp "random allocation").sh.) OR ((("allocat$" OR "assign$") ADJ2
"random$").ti,ab.) OR (("randomi#ed").ti,ab.) OR (("randomly").ti,ab.) OR
(("rct").ti,ab.) OR ((exp "placebos").sh.) OR (("placebo$").ti,ab.) OR ((exp
"prospective studies").pt.) OR ((exp "multicenter study").pt.) OR ((exp
"multicenter studies as topic").sh.) OR (("groups").ti,ab.) OR ((exp
"cross-over studies").sh.) OR (("crossover$" OR "cross over$" OR
"cross-over$").ti,ab.) OR ((exp "double blind method").sh.) OR ((exp "single
blind method").sh.) OR ((((("singl$" OR "doubl$" OR "treb$" OR "tripl$") ADJ1
("blind$" OR "mask$"))).ti,ab.)) NOT (((exp "animals").sh. NOT (exp
"humans").sh.) OR (("case report$").ti.) OR (("case reports").pt.) OR
(("letter").sh.) OR (("historical article").pt.) OR (("correspondence as
topic").sh.) OR (("meeting abstracts").pt.) OR (("abstracts").pt))) AND
(((("process$" ADJ "evaluat$").ti,ab.) NOT (("protocol").ti.)) AND ((("english"
OR "eng" OR "und" OR "undetermined").lg.) AND ((2012$ OR 2013$ OR 2014$ OR
2015$ OR 2016$).dp.) AND ((2011$ OR 2012$ OR 2013$ OR 2014$ OR 2015$ OR
2016$).dc.))))
+-----------------------------------------------------------------+
# Results
+-----------------------------------------------------------------+
Retrieved on 2017-02-27
571 hits
+-----------------------------------------------------------------+
```

B.5 Search engine: National Center for Biotechnology Information

B.5.1 Database: Pubmed

```
+-----------------------------------------------------------------+
# RCT filters
+-----------------------------------------------------------------+
## INCLUDED
("clinical trial"[pt])
("clinical trials as topic"[mh])
("clinical trial"[tiab] OR "clinical trials"[tiab])
(trial*[tiab])
(drug therapy[sh])

("random allocation"[mh])
(allocat*[tiab] AND random*[tiab])
(randomized[tiab] OR randomised[tiab])
(randomly[tiab])
(rct[tiab])

("placebos"[mh])
(placebo*[tiab])

("prospective studies"[mh])

("cross-over studies"[mh])
("crossover"[tiab] OR "cross over"[tiab] OR "cross-over"[tiab])

("multicenter study"[pt])
("multicenter studies as topic"[mh])
(groups[tiab])

("double blind method"[mh])
("single blind method"[mh])
((singl*[tiab] OR doubl*[tiab] OR tripl*[tiab] OR trebl*[tiab]) AND
(blind*[tiab] OR mask*[tiab]))
## EXCLUDED
("animals"[mh] NOT "humans"[mh])
("case report"[ti])
("case reports"[ti])
```

```
("case reports"[pt])
("letter"[pt])
("historical article"[pt])
("correspondence as topic"[mh])
+-------------------------------------------------------------------+
# Subject filters
+-------------------------------------------------------------------+
## INCLUDED
("process evaluation"[tiab])
("process evaluations"[tiab])
("process evaluator"[tiab])
("process evaluators"[tiab])
## EXCLUDED
("protocol"[ti])
## LIMITS (Combined with AND)
(eng[la] OR und[la])

("2012/01/01"[pdat] : "2016/12/31"[pdat])
("2011/01/01"[crdat] : "2016/12/31"[crdat])

* Reference for dates: https://www.ncbi.nlm.nih.gov/books/NBK179288/
[PDAT]  Date - Publication
[CRDT]  Date - Create
+-------------------------------------------------------------------+
# Combined
+-------------------------------------------------------------------+
((((("clinical trial"[pt]) OR ("clinical trials as topic"[mh]) OR ("clinical
trial"[tiab] OR "clinical trials"[tiab]) OR (trial*[tiab]) OR (drug
therapy[sh]) OR ("random allocation"[mh]) OR (allocat*[tiab] AND random*[tiab])
OR (randomized[tiab] OR randomised[tiab]) OR (randomly[tiab]) OR (rct[tiab]) OR
("placebos"[mh]) OR (placebo*[tiab]) OR ("prospective studies"[mh]) OR
("cross-over studies"[mh]) OR ("crossover"[tiab] OR "cross over"[tiab] OR
"cross-over"[tiab]) OR ("multicenter study"[pt]) OR ("multicenter studies as
topic"[mh]) OR (groups[tiab]) OR ("double blind method"[mh]) OR ("single blind
method"[mh]) OR ((singl*[tiab] OR doubl*[tiab] OR tripl*[tiab] OR trebl*[tiab])
AND (blind*[tiab] OR mask*[tiab]))) NOT (("animals"[mh] NOT "humans"[mh]) OR
("case report"[ti]) OR ("case reports"[ti]) OR ("case reports"[pt]) OR
("letter"[pt]) OR ("historical article"[pt]) OR ("correspondence as
topic"[mh]))) AND ((((("process evaluation"[tiab]) OR ("process
evaluations"[tiab]) OR ("process evaluator"[tiab]) OR ("process
evaluators"[tiab])) NOT ("protocol"[ti])) AND ((eng[la] OR und[la]) AND
("2012/01/01"[pdat] : "2016/12/31"[pdat]) AND ("2011/01/01"[crdat] :
"2016/12/31"[crdat]))))
+-------------------------------------------------------------------+
# Results
```

```
+----------------------------------------------------------------------+
Retrieved on 2017-02-27
580 hits
+----------------------------------------------------------------------+
```

Appendix C

Second reviewer's manual

Manual for conducting secondary review of a systematic review on methods used in process evaluation

Li, Ruijie

September 2017

Contents

1 Aims of systematic review

The aim of the systematic review is to understand the aims, meaning and methods of process evaluation in mental health studies. This is to facilitate the identification of gaps in the application of current statistical methods for the purpose of conducting causal inferences. This systematic review is part of a PhD project titled "Statistical Methods for Process Evaluation in Randomised Clinical Trials".

Briefly, process evaluation (PE) is a form of evaluation often conducted alongside outcomes evaluation (OE). While OE seeks to establish the efficacy or effectiveness of an intervention, PE seeks to understand the variability in the processes that go into the conduct of the intervention. These processes can be characteristics of the person providing the intervention, the intervention itself, the patient or the context in which the intervention is conducted in. Understanding the variability in these processes can provide an understanding of the potential causal mechanisms of the intervention.

1.1 Methods for systematic review

The following are the steps taken to derive the final set of articles for review.

1. Identify the database and search engines to use based on scope of review.
2. Identify keywords from the literature.
   (a) Two sets of keywords for the identification of randomised controlled trials (RCTs) and PE studies respectively.
   (b) RCT keywords were derived from Cochrane and Scottish Intercollegiate Guidelines Network (SIGN) and modified for use in the different search engines and databases.
   (c) PE keywords were initially identified from literature and its variants were identified from a keyword study to ensure adequate coverage of the concept of PE.
3. Screen the articles according to the inclusion and exclusion criteria
   (a) The screening was split into two steps. The first step was to identify papers that had a primary trial conducted using the RCT design and a PE.
   (b) The second step was to further differentiate between papers that had a mental health focus and those that do not.
4. Enter the information from the identified papers into the database.

2 Aims of secondary review

The systematic review was conducted in such a way to enable reproducibility. However, there exist parts that have an element of subjectivity and the aim of the secondary review is to ensure that these steps are reproducible with the use of a set of guidelines. Therefore, the aim of the secondary review is to reproduce parts steps 3 and 4 of the methodology.

2.1 Methods for secondary review

For reproducing the screening step of the systematic review, the two screening steps will be merged. The inclusion and exclusion criteria can be found in Appendix B. The secondary reviewer will screen from a set of 30 articles to identify those that meet the criteria. The identified articles will then be entered in to the database using the guide found in Appendix C. After entering the data into the database, the second reviewer will be provided with the data entered by the first reviewer. The second reviewer will then highlight any significant discrepancies.

3 Tasks for secondary review

1. Screen from a set of 30 articles to identify suitable articles.
2. From the identified articles, enter the information into the database.

Glossary

causal inference

The process by which causality is inferred. In the context of this document, causal inference refers to any attempts to infer causality.

efficacy and effectiveness

There are multiple definitions for efficacy and effectiveness in the literature. They mostly refer to the same underlying concept with different orientations. Both efficacy and effectiveness concerns with whether an intervention produces the anticipated results. One definition considers the efficacy to be whether the results are produced under ideal conditions and effectiveness to be whether it produces the results under real world conditions (Gartlehner, Hansen, Nissman, Lohr & Carey, 2006). Another definition considers efficacy to be the results assessed when the intervention is offered and effectiveness to be the results assessed when the intervention is received (Dunn et al., 2015). The two definitions are similar with the latter being more precise about the intervention effects being measured.

mediation

The process by which the cause and effect are related indirectly, via an alternate mechanism.

moderation

The process by which the intervention effects differ amongst people by their baseline characteristics (Kraemer, Wilson, Fairburn & Agras, 2002).

outcomes evaluation (OE)

OE is a form of evaluation used to assess the efficacy or effectiveness of an intervention. This is commonly employed in clinical trials and programme evaluations. The RCT design is commonly used as it provides a rigorous framework to infer causality through the use of counterfactuals.

post-randomisation treatment effect modification (PREM)

Similar to and often confused with moderation. While moderation refers to the difference in treatment effects according to baseline characteristics, PREM refers to the difference in treatment effects in characteristics that are not measured at baseline. The analysis of PREM differs from a moderation analysis and a common way to conduct such an analysis involve the use of an instrumental variable (Dunn et al., 2015).

process evaluation (PE)

PE is a form of evaluation commonly used to assess the processes involved in delivering an intervention. This form of evaluation makes a comparison between the intent or theory underlying the intervention with what is borne out in reality. A common purpose for conducting such an evaluation is to understand the effects of the processes on the outcomes of interest. More information on PE can be found in a recent Medical Research Council (MRC) guideline on the use of PE for complex interventions (Moore et al., 2015)

A Materials for secondary review

A set of bibliographic data together with the full text of the papers will be provided to the second reviewer. These will be in the form of a Zotero database that comes with several searching and tagging functions to facilitate the screening of the papers. A Microsoft Excel database will be provided for the entering of the data. The definitions of the fields for the database is included in the subsequent appendix.

B Inclusion and exclusion criteria

Inclusion criteria:

1. peer-reviewed journal article
2. has an intervention that is assessed in an OE using an RCT design.
3. the PE can be conducted using any design.
4. the primary outcome has relevance to mental health (e.g. Patient Health Questionnaire (PHQ-9) scores for depression, mental health scores from quality of life instruments.).
5. the target population need not necessarily be for people with mental health disorders.
6. studies that assess interventions and have a primary mental health outcome

Exclusion criteria:

1. non-peer reviewed publications (e.g. dissertation, protocol, poster, conference paper)
2. studies assessing interventions for people with mental health disorders without a primary mental health outcome (e.g. physical activity intervention for elderly with dementia with the 6 minute walk test as a primary outcome).
3. studies that do not have a primary mental health related outcome or have mental health related outcomes as secondary outcomes only.

C Field definitions for database

Table 1: Field definitions

Field	Definition
Study metadata:	
Publication	publication the paper was published in
Year	year paper was published
Author	authors of the paper
Title	title of the paper
Status	accept/reject: status after screening
Outcomes evaluation:	
Main trial outcome	positive/negative
Design:	
type	parallel/others: RCT design
arms	number of arms
blinding	presence and type of blinding
Randomisation:	
type	simple/cluster/stratified
level	subject/cluster
Subjects:	
who (condition)	target population characteristic
who (age)	average age of subjects
where (country)	country study was conducted in
where (facility)	type of facility where subjects come from
how (recruited)	mechanism by which subjects were recruited
number (total)	total number of subjects
number (arms)	number of subjects in each arm
Outcomes:	
(aim)	efficacy/effectiveness: for studies that state their aim in terms of efficacy or effectiveness, the stated aim will be entered unless contrary information exists in the paper strongly suggesting the opposite (e.g. a study with the stated aim of assessing efficacy of an intervention in a unique context when prior studies have already shown it to be efficacious. For a detailed discussion on the differences, refer to Gartlehner et al. (2006).
what (primary)	the primary outcome construct (e.g. depression, anxiety)
what (measures)	the primary outcome measure
when	at what time points were the primary outcomes measured
what (analysis)	what analysis was conducted
Intervention:	

Table 1: (continued)

Field	Definition
treatment	the intervention under evaluation
control	the control intervention
Process evaluation:	
Design:	
type	selected arms of trial only (which arm?)/all arms of trial
paradigm	quantitative/qualitative/mixed
framework	PE framework/not stated
aim (hypothesis)	hypothesis generating/confirming
aim (what to evaluate)	what is the aim of the PE
when (relative to trial)	before/alongside/after: when was the evaluation conducted in relation to the main trial
when (assessment time points)	e.g. baseline, 1 month, 2 months: at what time points was the evaluation conducted
link with OE	yes/no: were the results of the PE analysed in the context of the OE results?
Subjects:	
who	e.g. intervention group, staff: Who were included in the study? If different for the quantitative and qualitative component, state who were involved in each.
how (recruited)	e.g. patients at the clinic were invited: how were the subjects recruited to the study
Quantitative:	
Subjects	
number (total)	total number of subjects
sampling	e.g. same as OE, convenience sample: how were the subjects sampled?
Process	
what (construct)	e.g. dose, adherence: what is the process that is being evaluated?
what (analysis)	e.g. descriptive statistics, ANOVA: what methods were used for the analysis?
causal inference aim	yes/no: was casual inference a stated aim?
causal inference implied	yes/no: was causal inference implied as an aim?
causal inference type	e.g. mediation, moderation: what form of casual inference was stated/implied?
conclusion justified?	yes/no: were the conclusions justified?
conclusion justified (why)	e.g. the analytic methods provided support for the claims made in the conclusion: explain why the conclusion are justified/not justified.

Table 1: (continued)

Field	Definition
measurement error	yes/no: was measurement error accounted for in the analyses?
mediation	yes/no: was any form of mediation analyses conducted?
PREM	yes/no: was any form of post randomisation effect modification analysis conducted?
Qualitative:	
Subjects	
number (total)	the total number of subjects
sampling	e.g. convenience/purposive for maximum variability: Can be a combination of several ways of sampling. If purposive, include any stated purpose of sampling.
Process	
what	e.g. patient satisfaction: what is the process that is being evaluated?
what	e.g. content analysis: what methods/qualitative framework is being used to analyse the data?
Remarks	Any other remarks

References

Dunn, G., Emsley, R., Liu, H., Landau, S., Green, J., White, I. & Pickles, A. (2015). Evaluation and validation of social and psychological markers in randomised trials of complex interventions in mental health: A methodological research programme. *Health technology assessment (Winchester, England)*, *19*(93), 1–116. doi:10.3310/hta19930

Gartlehner, G., Hansen, R. A., Nissman, D., Lohr, K. N. & Carey, T. S. (2006). *Criteria for Distinguishing Effectiveness From Efficacy Trials in Systematic Reviews*. Agency for Healthcare Research and Quality (US). Rockville (MD). Retrieved April 11, 2016, from http://www.ncbi.nlm.nih.gov/books/NBK44029/

Kraemer, H. C., Wilson, G. T., Fairburn, C. G. & Agras, W. S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of general psychiatry*, *59*(10), 877–883. Retrieved November 4, 2016, from http://archpsyc.jamanetwork.com/article.aspx?articleid=206766

Moore, G., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., ... Baird, J. (2015). *Process evaluation of complex interventions: UK Medical Research Council (MRC) guidance*.

Appendix D

Systematic review results: Population and treatment

Table D.1: Accepted studies trial population and treatment

No.	Study	Target population	Intervention
01.	01.	people with advanced lung cancer who require radiotherapy	psycho-educational intervention for symptom management
02.	02.	people with post-traumatic stress disorder, generalized anxiety disorder, panic disorder or social anxiety disorder	**CALM** (anxiety management and cognitive behavioural therapy components) and medication optimisation
03.	03.	women with post-partum depression	psycho-education derived from interpersonal psychotherapy principles
04.	04.	people residing in nursing homes with depression	multidisciplinary, evidence based care program to improve the management of depression in nursing home residents
05.	05.	nurses and allied health professionals with mental health issues	online screening for mental health issues and appointment with occupational physician; online screening for mental health issues with digital content to improve
06.	06.	people with schizophrenia	activity group; art therapy
07.	07.	students	attention control; classroom-based **CBT**
07.	08.	students	attention control; classroom-based **CBT**
08.	09.	people with stroke and aphasia	behavioural therapy
09.	10.	people with psychosis	web-based tool to help patients decide what treatment modalities they would like

Table D.1: Accepted studies trial population and treatment (cont'd)

No.	Study	Target population	Intervention
10.	11.	workers with common mental disorders	common mental disorders management training for occupational physicians
11.	12.	people residing in care homes	exercise
12.	13.	people with fear of falling	cognitive behavioural therapy
13.	14.	workers with depressive symptoms	web-based problem-solving treatment and cognitive therapy
14.	15.	older adults residing in the community who are at risk of social isolation	telephone calls from volunteers to befriend subjects
15.	16.	people with psychosis	support patients in recovery after psychosis
15.	17.	people with psychosis	support patients in recovery after psychosis
16.	18.	people with eating disorders	cognitive interpersonal treatment
16.	19.	people with eating disorders	cognitive interpersonal treatment
16.	20.	people with eating disorders	cognitive interpersonal treatment
17.	21.	carers of young people with a first episode psychosis	bibliotherapy
18.	22.	people with HIV/AIDS who are depressed	group based interpersonal therapy
19.	23.	people with dementia and their caregivers	home-based physical exercise training, psycho-education, communication skills training, and pleasant activities training
20.	24.	women at risk of stress, anxiety, depression	mindfulness based group therapy
21.	25.	people with cancer-related fatigue	web-based resource

Table D.1: Accepted studies trial population and treatment (cont'd)

No.	Study	Target population	Intervention
21.	26.	people with cancer-related fatigue	web-based resource
22.	27.	women who are first time mothers with an uncomplicated pregnancy	postnatal psycho-education
23.	28.	primary school children	manualised **CBT** programme
24.	29.	school children with hyperactivity/inattention	management strategies of children with **ADHD** provided for the parents and/or teachers
25.	30.	people with schizophrenia and associated disorders	assessments based on structured patient-clinician dialogue and solution focused therapy

Appendix E

Systematic review: process evaluation studies identified

Arends, I., Bültmann, U., Nielsen, K., van Rhenen, W., de Boer, M. R., & van der Klink, J. J. L. (2014). Process evaluation of a problem solving intervention to prevent recurrent sickness absence in workers with common mental disorders. *Social Science & Medicine (1982)*, *100*, 123–132. https://doi.org/10.1016/j.socscimed.2013.10.041

Chan, C. W. H., Richardson, A., & Richardson, J. (2012). Evaluating a complex intervention: A process evaluation of a psycho-education program for lung cancer patients receiving palliative radiotherapy. *Special Issue: Advances in contemporary complex health care: Nursing interventions.*, *40*(2), 234–244. https://doi.org/10.5172/conu.2012.40.2.234

Curran, G. M., Sullivan, G., Mendel, P., Craske, M. G., Sherbourne, C. D., Stein, M. B., McDaniel, A., & Roy-Byrne, P. (2012). Implementation of the CALM intervention for anxiety disorders: A qualitative study. *Implementation science: IS*, *7*, 1–11. https://doi.org/10.1186/1748-5908-7-14

Ellard, D. R., Thorogood, M., Underwood, M., Seale, C., & Taylor, S. J. C. (2014). Whole home exercise intervention for depression in older care home residents (the OPERA study): A process evaluation. *BMC medicine*, *12*, 1. https://doi.org/10.1186/1741-7015-12-1

Finch, T. L., Bamford, C., Deary, V., Sabin, N., & Parry, S. W. (2014). Making sense of a cognitive behavioural therapy intervention for fear of falling: Qualitative study of intervention

development. *BMC health services research*, *14*, 436. https://doi.org/10.1186/1472-6963-14-436

Foster, C., Grimmett, C., May, C. M., Ewings, S., Myall, M., Hulme, C., Smith, P. W., Powers, C., Calman, L., Armes, J., Breckons, M., Corner, J., Fenlon, D., Batehup, L., Lennan, E., R May, C., Morris, C., Neylon, A., Ream, E., … Richardson, A. (2016). A web-based intervention (RESTORE) to support self-management of cancer-related fatigue following primary cancer treatment: A multi-centre proof of concept randomised controlled trial. *Supportive Care in Cancer: Official Journal of the Multinational Association of Supportive Care in Cancer*, *24*(6), 2445–2453. https://doi.org/10.1007/s00520-015-3044-7

Gao, L.-l., Luo, S.-y., & Chan, S. W.-c. (2012). Interpersonal psychotherapy-oriented program for Chinese pregnant women: Delivery, content, and personal impact. *Nursing & Health Sciences*, *14*(3), 318–324. https://doi.org/10.1111/j.1442-2018.2012.00722.x

Geraedts, A. S., Kleiboer, A., Wiezer, N., Cuijpers, P., van Mechelen, W., & Anema, J. (2014). Feasibility of a worker-directed web-based intervention for employees with depressive symptoms. *Internet Interventions*, *1*(3), 132–140. https://doi.org/10.1016/j.invent.2014.07.001

Hind, D., Mountain, G., Gossage-Worrall, R., Walters, S. J., Duncan, R., Newbould, L., Rex, S., Jones, C., Bowling, A., Cattan, M., Cairns, A., Cooper, C., Goyder, E., & Edwards, R. T. (2014). Putting Life in Years (PLINY): A randomised controlled trial and mixed-methods process evaluation of a telephone friendship intervention to improve mental well-being in independently living older people. *Public Health Research*, *2*(7), 1–222. https://doi.org/10.3310/phr02070

Ketelaar, S. M., Gärtner, F. R., Bolier, L., Smeets, O., Nieuwenhuijsen, K., & Sluiter, J. K. (2013). Mental Vitality @ Work–a workers' health surveillance mental module for nurses and allied health care professionals: Process evaluation of a randomized controlled trial. *Journal of Occupational and Environmental Medicine*, *55*(5), 563–571. https://doi.org/10.1097/JOM.0b013e318289ee3e

Leamy, M., Clarke, E., Le Boutillier, C., Bird, V., Janosik, M., Sabas, K., Riley, G., Williams, J., & Slade, M. (2014). Implementing a complex intervention to support personal recovery: A qualitative study nested within a cluster randomised controlled trial. *PloS One*, *9*(5), e97091. https://doi.org/10.1371/journal.pone.0097091

Leontjevas, R., Gerritsen, D. L., Koopmans, R. T. C. M., Smalbrugge, M., & Vernooij-Dassen, M. J. F. J. (2012). Process evaluation to explore internal and external validity of the 'Act in Case of Depression' care program in nursing homes. *Journal of the American Medical Directors Association*, *13*(5), 488.e1–8. https://doi.org/10.1016/j.jamda.2012.03.006

Lose, A., Davies, C., Renwick, B., Kenyon, M., Treasure, J., Schmidt, U., & MOSAIC trial group. (2014). Process evaluation of the maudsley model for treatment of adults with anorexia nervosa trial. Part II: Patient experiences of two psychological therapies for treatment of anorexia nervosa. *European Eating Disorders Review: The Journal of the Eating Disorders Association*, *22*(2), 131–139. https://doi.org/10.1002/erv.2279

McCann, T. V., & Lubman, D. I. (2014). Qualitative process evaluation of a problem-solving guided self-help manual for family carers of young people with first-episode psychosis. *BMC psychiatry*, *14*, 168. https://doi.org/10.1186/1471-244X-14-168

Myall, M., May, C. R., Grimmett, C., May, C. M., Calman, L., Richardson, A., & Foster, C. L. (2015). RESTORE: An exploratory trial of a web-based intervention to enhance self-management of cancer-related fatigue: Findings from a qualitative process evaluation. *BMC medical informatics and decision making*, *15*, 94. https://doi.org/10.1186/s12911-015-0214-y

Omer, S., Golden, E., & Priebe, S. (2016). Exploring the Mechanisms of a Patient-Centred Assessment with a Solution Focused Approach (DIALOG+) in the Community Treatment of Patients with Psychosis: A Process Evaluation within a Cluster-Randomised Controlled Trial. *PloS One*, *11*(2), e0148415. https://doi.org/10.1371/journal.pone.0148415

Patterson, S., Borschmann, R., & Waller, D. E. (2013). Considering referral to art therapy: Responses to referral and experiences of participants in a randomised controlled trial. *International Journal of Art Therapy*, *18*(1), 2–9. https://doi.org/10.1080/17454832.2012.738425

Petersen, I., Hanass Hancock, J., Bhana, A., & Govender, K. (2014). A group-based counselling intervention for depression comorbid with HIV/AIDS using a task shifting approach in South Africa: A randomized controlled pilot study. *Journal of Affective Disorders*, *158*, 78–84. https://doi.org/10.1016/j.jad.2014.02.013

Prick, A.-E., de Lange, J., van 't Leven, N., & Pot, A. M. (2014). Process evaluation of a multicomponent dyadic intervention study with exercise and support for people with dementia and their family caregivers. *Trials*, *15*, 401. https://doi.org/10.1186/1745-6215-15-401

Shorey, S., Chan, S. W.-C., Chong, Y. S., & He, H.-G. (2015). Perceptions of primiparas on a postnatal psychoeducation programme: The process evaluation. *Midwifery*, *31*(1), 155–163. https://doi.org/10.1016/j.midw.2014.08.001

Stallard, P., Phillips, R., Montgomery, A. A., Spears, M., Anderson, R., Taylor, J., Araya, R., Lewis, G., Ukoumunne, O. C., Millings, A., Georgiou, L., Cook, E., & Sayal, K. (2013). A cluster randomised controlled trial to determine the clinical effectiveness and cost-effectiveness of classroom-based cognitive-behavioural therapy (CBT) in reducing symptoms of depression in high-risk adolescents. *Health Technology Assessment*, *17*(47), vii–xvii, 1–109. https://doi.org/10.3310/hta17470

Stallard, P., Skryabina, E., Taylor, G., Anderson, R., Ukoumunne, O. C., Daniels, H., Phillips, R., & Simpson, N. (2015). A cluster randomised controlled trial comparing the effectiveness and cost-effectiveness of a school-based cognitive-behavioural therapy programme (FRIENDS) in the reduction of anxiety and improvement in mood in children aged 9/10 years. *Public Health Research*, *3*(14), 1–88. https://doi.org/10.3310/phr03140

Taylor, J. A., Phillips, R., Cook, E., Georgiou, L., Stallard, P., & Sayal, K. (2014). A qualitative process evaluation of classroom-based cognitive behaviour therapy to reduce adolescent depression. *International Journal of Environmental Research and Public Health*, *11*(6), 5951–5969. https://doi.org/10.3390/ijerph110605951

Taylor, J. A., Valentine, A. Z., Sellman, E., Bransby-Adams, K., Daley, D., & Sayal, K. (2015). A qualitative process evaluation of a randomised controlled trial of a parenting intervention in community (school) settings for children at risk of attention deficit hyperactivity disorder (ADHD). *BMC psychiatry*, *15*, 290. https://doi.org/10.1186/s12888-015-0670-z

Thomas, S. A., Russell, C., Seed, R., Worthington, E., Walker, M. F., Macniven, J. A., & Lincoln, N. B. (2013). An evaluation of treatment integrity in a randomized trial of behavioural therapy for low mood in stroke patients with aphasia. *Clinical Rehabilitation*, *27*(12), 1097–1106. https://doi.org/10.1177/0269215513489579

van der Krieke, L., Emerencia, A. C., Boonstra, N., Wunderink, L., de Jonge, P., & Sytema, S. (2013). A web-based tool to support shared decision making for people with a psychotic disorder: Randomized controlled trial and process evaluation. *Journal of Medical Internet Research*, *15*(10), e216. https://doi.org/10.2196/jmir.2851

Wallace, G., Bird, V., Leamy, M., Bacon, F., Le Boutillier, C., Janosik, M., MacPherson, R., Williams, J., & Slade, M. (2016). Service user experiences of REFOCUS: A process evaluation of a

pro-recovery complex intervention. *Social Psychiatry and Psychiatric Epidemiology*, *51*(9), 1275–1284. https://doi.org/10.1007/s00127-016-1257-9

Waterman-Collins, D., Renwick, B., Lose, A., Kenyon, M., Serpell, L., Richards, L., Boughton, N., Treasure, J., Schmidt, U., & MOSAIC trial group. (2014). Process evaluation of the MOSAIC Trial, Part I: Therapist experiences of delivering two psychological therapies for treatment of anorexia nervosa. *European Eating Disorders Review: The Journal of the Eating Disorders Association*, *22*(2), 122–130. https://doi.org/10.1002/erv.2278

Woolhouse, H., Mercuri, K., Judd, F., & Brown, S. J. (2014). Antenatal mindfulness intervention to reduce depression, anxiety and stress: A pilot randomised controlled trial of the MindBabyBody program in an Australian tertiary maternity hospital. *BMC pregnancy and childbirth*, *14*, 369. https://doi.org/10.1186/s12884-014-0369-z

Zainal, K. A., Renwick, B., Keyes, A., Lose, A., Kenyon, M., DeJong, H., Broadbent, H., Serpell, L., Richards, L., Johnson-Sabine, E., Boughton, N., Whitehead, L., Treasure, J., Schmidt, U., & MOSAIC trial group. (2016). Process evaluation of the MOSAIC trial: Treatment experience of two psychological therapies for out-patient treatment of Anorexia Nervosa. *Journal of Eating Disorders*, *4*, 2. https://doi.org/10.1186/s40337-016-0091-5

Appendix F

Systematic review: Primary studies of process evaluation

Arends, I., Klink, J. J. L. van der, Rhenen, W. van, Boer, M. R. de, & Bültmann, U. (2014). Prevention of recurrent sickness absence in workers with common mental disorders: Results of a cluster-randomised controlled trial. *Occup Environ Med*, *71*(1), 21–29. https://doi.org/10.1136/oemed-2013-101412

Chan, C. W. H., Richardson, A., & Richardson, J. (2011). Managing Symptoms in Patients with Advanced Lung Cancer During Radiotherapy: Results of a Psychoeducational Randomized Controlled Trial. *Journal of Pain and Symptom Management*, *41*(2), 347–357. https://doi.org/10.1016/j.jpainsymman.2010.04.024

Crawford, M. J., Killaspy, H., Barnes, T. R. E., Barrett, B., Byford, S., Clayton, K., Dinsmore, J., Floyd, S., Hoadley, A., Johnson, T., Kalaitzaki, E., King, M., Leurent, B., Maratos, A., O'Neill, F. A., Osborn, D. P., Patterson, S., Soteriou, T., Tyrer, P., & Waller, D. (2012). Group art therapy as an adjunctive treatment for people with schizophrenia: Multicentre pragmatic randomised trial. *BMJ*, *344*, e846. https://doi.org/10.1136/bmj.e846

Gao, L.-l., Chan, S. W.-c., Li, X., Chen, S., & Hao, Y. (2010). Evaluation of an interpersonal-psychotherapy-oriented childbirth education programme for Chinese first-time childbearing women: A randomised controlled trial. *International Journal of Nursing Studies*, *47*(10), 1208–1216. https://doi.org/10.1016/j.ijnurstu.2010.03.002

Gärtner, F. R., Nieuwenhuijsen, K., Ketelaar, S. M., Dijk, F. J. h van, & Sluiter, J. K. (2013). The Mental Vitality @ Work Study: Effectiveness of a Mental Module for Workers' Health Surveillance for Nurses and Allied Health Care Professionals on Their Help-seeking Behavior. *Journal of Occupational and Environmental Medicine*, *55*(10), 1219–1229. https://doi.org/10.1097/JOM.0b013e31829f310a

Geraedts, A. S., Kleiboer, A. M., Twisk, J., Wiezer, N. M., Mechelen, W. van, & Cuijpers, P. (2014). Long-Term Results of a Web-Based Guided Self-Help Intervention for Employees With Depressive Symptoms: Randomized Controlled Trial. *Journal of Medical Internet Research*, *16*(7), e168. https://doi.org/10.2196/jmir.3539

Leontjevas, R., Gerritsen, D. L., Smalbrugge, M., Teerenstra, S., Vernooij-Dassen, M. J., & Koopmans, R. T. (2013). A structural multidisciplinary approach to depression management in nursing-home residents: A multicentre, stepped-wedge cluster-randomised trial. *The Lancet*, *381*(9885), 2255–2264. https://doi.org/10.1016/S0140-6736(13)60590-5

McCann, T. V., Lubman, D. I., Cotton, S. M., Murphy, B., Crisp, K., Catania, L., Marck, C., & Gleeson, J. F. M. (2013). A Randomized Controlled Trial of Bibliotherapy for Carers of Young People With First-Episode Psychosis. *Schizophrenia Bulletin*, *39*(6), 1307–1317. https://doi.org/10.1093/schbul/sbs121

Parry, S. W., Bamford, C., Deary, V., Finch, T. L., Gray, J., MacDonald, C., McMeekin, P., Sabin, N. J., Steen, I. N., Whitney, S. L., & McColl, E. M. (2016). Cognitive-behavioural therapy-based intervention to reduce fear of falling in older people: Therapy development and randomised controlled trial - the Strategies for Increasing Independence, Confidence and Energy (STRIDE) study. *Health Technology Assessment (Winchester, England)*, *20*(56), 1–206. https://doi.org/10.3310/hta20560

Prick, A.-E., de Lange, J., Scherder, E., Twisk, J., & Pot, A. M. (2016). The effects of a multicomponent dyadic intervention on the mood, behavior, and physical health of people with dementia: A randomized controlled trial. *Clinical Interventions in Aging*, 383. https://doi.org/10.2147/CIA.S95789

Priebe, S., Kelley, L., Omer, S., Golden, E., Walsh, S., Khanom, H., Kingdon, D., Rutterford, C., McCrone, P., & McCabe, R. (2015). The Effectiveness of a Patient-Centred Assessment with a Solution-Focused Approach (DIALOG+) for Patients with Psychosis: A Pragmatic Cluster-Randomised Controlled Trial in Community Care. *Psychotherapy and Psychosomatics*, *84*(5), 304–313. https://doi.org/10.1159/000430991

Roy-Byrne, P., Craske, M. G., Sullivan, G., Rose, R. D., Edlund, M. J., Lang, A. J., Bystritsky, A., Welch, S. S., Chavira, D. A., Golinelli, D., Campbell-Sills, L., Sherbourne, C. D., & Stein, M. B. (2010). Delivery of Evidence-Based Treatment for Multiple Anxiety Disorders in Primary Care: A Randomized Controlled Trial. *JAMA*, *303*(19), 1921–1928. https://doi.org/10.1001/jama.2010.608

Sayal, K., Taylor, J. A., Valentine, A., Guo, B., Sampson, C. J., Sellman, E., James, M., Hollis, C., & Daley, D. (2016). Effectiveness and cost-effectiveness of a brief school-based group programme for parents of children at risk of ADHD: A cluster randomised controlled trial. *Child: Care, Health and Development*, *42*(4), 521–533. https://doi.org/10.1111/cch.12349

Schmidt, U., Magill, N., Renwick, B., Keyes, A., Kenyon, M., Dejong, H., Lose, A., Broadbent, H., Loomes, R., Yasin, H., Watson, C., Ghelani, S., Bonin, E.-M., Serpell, L., Richards, L., Johnson-Sabine, E., Boughton, N., Whitehead, L., Beecham, J., … Landau, S. (2015). The Maudsley Outpatient Study of Treatments for Anorexia Nervosa and Related Conditions (MOSAIC): Comparison of the Maudsley Model of Anorexia Nervosa Treatment for Adults (MANTRA) with specialist supportive clinical management (SSCM) in outpatients with broadly defined anorexia nervosa: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *83*(4), 796–807. https://doi.org/10.1037/ccp0000019

Shorey, S., Chan, S. W. C., Chong, Y. S., & He, H.-G. (2015). A randomized controlled trial of the effectiveness of a postnatal psychoeducation programme on self-efficacy, social support and postnatal depression among primiparas. *Journal of Advanced Nursing*, *71*(6), 1260–1273. https://doi.org/10.1111/jan.12590

Slade, M., Bird, V., Clarke, E., Boutillier, C. L., McCrone, P., Macpherson, R., Pesola, F., Wallace, G., Williams, J., & Leamy, M. (2015). Supporting recovery in patients with psychosis through care by community-based adult mental health teams (REFOCUS): A multisite, cluster, randomised, controlled trial. *The Lancet Psychiatry*, *2*(6), 503–514. https://doi.org/10.1016/S2215-0366(15)00086-3

Thomas, S. A., Walker, M. F., Macniven, J. A., Haworth, H., & Lincoln, N. B. (2013). Communication and Low Mood (CALM): A randomized controlled trial of behavioural therapy for stroke patients with aphasia. *Clinical Rehabilitation*, *27*(5), 398–408. https://doi.org/10.1177/0269215512462227

Underwood, M., Lamb, S. E., Eldridge, S., Sheehan, B., Slowther, A.-M., Spencer, A., Thorogood, M., Atherton, N., Bremner, S. A., Devine, A., Diaz-Ordaz, K., Ellard, D. R., Potter, R.,

Spanjers, K., & Taylor, S. J. (2013). Exercise for depression in elderly residents of care homes: A cluster-randomised controlled trial. *The Lancet*, *382*(9886), 41–49. https://doi.org/10.1016/S0140-6736(13)60649-2

Appendix G

Computer programme: Model- and simulation-based estimators

G.1 Auxiliary functions

The *gen.utilities* function generates the various auxiliary functions used in the main programme. Use the command *list2env(gen.utilities())* to insert these functions into the present global environment.

```
gen.utilities <- function() {
# functions
## utilities
    ut.box <- function(v.contents, chr.head = "┼", chr.sep = "─",
                       chr.tail = "┼", chr.wall = "│", s.align = "l") {
        v.box <- c(lh = "─",
                   ctl = "┌", ctr = "┐", cbl = "└", cbr = "┘",
                   lvhl = "├", lvhr = "┤", lhvt = "┬", lhvb = "┴",
                   lc = "┼")
        if(class(v.contents) |>
           match(c("integer", "numeric", "character"),
                 nomatch = 0) |>
           (any |> Negate())())
            stop("Incorrect contents format.")

        s.contents.lenmax <- nchar(v.contents) |> max()
```

```
    s.box.lenmax <- length(c(chr.head, chr.sep, chr.tail)) |> sum()
    if(s.contents.lenmax < s.box.lenmax)
        s.contents.lenmax <- s.box.lenmax
    s.contents.lenpadded <- s.contents.lenmax + 2
    s.contents.lenpadded <- (4 - (s.contents.lenpadded %% 4)) + s.contents.lenpadded

    s.contents.lenpadded.use <-
        switch(s.align,
               l = -s.contents.lenpadded,
               r = s.contents.lenpadded,
               c = s.contents.lenpadded/2)
    v.messages <-
        switch((s.align == "c") + 1,
               sprintf(paste0(chr.wall, " %*s ", chr.wall),
                       s.contents.lenpadded.use, v.contents),
               {
                   l.contents <-
                       Map(function(x)
                           {
                               s.len <- length(x) %/% 2
                               Map(function(v) paste0(v, collapse = ""),
                                   list(head(x, s.len), tail(x, -s.len)))
                           }, strsplit(v.contents, ""))

                   mapply(function(x)
                          {
                              paste0(sprintf(paste0(chr.wall, " %*s"),
                                             s.contents.lenpadded.use, x[[1]]),
                                     sprintf(paste0("%*s ", chr.wall),
                                             -s.contents.lenpadded.use, x[[2]]),
                                     collapse = "")
                          }, l.contents)
               })
    v.sep <- paste0(c(chr.head,
                      rep(chr.sep,
                          s.contents.lenpadded +
                              ((nchar(c(" ", chr.wall)) * 2) |> sum()) -
                              (nchar(c(chr.head, chr.tail))  |> sum())),
                      chr.tail),
                    collapse = "")
    mapply(message, c(v.sep, v.messages, v.sep)) |> invisible()
}
ut.clear.disp <- function() {
    if(!is.null(dev.list())) {
        invisible(mapply(dev.off, dev.list()))
```

```
    }
    switch(commandArgs()[1],
           RStudio = cat("\014"),
           system("clear"))
}
ut.clear.dispdev <- function() {
if(!is.null(dev.list()))
    ut.silent(mapply(dev.off, dev.list()))
}
ut.clear.globalwspc <- function() {
    local(rm(list = ls()), envir = .GlobalEnv)
}
ut.get.all <- function() {
    env <- parent.frame(1)
    mget(ls(envir = env), envir = env)
}
ut.suppress <- function(...) {
    suppressMessages(...) |> suppressWarnings()
}
ut.silent <- function(...) {
    sink(nullfile())
    try(..., silent = TRUE, outFile = nullfile()) |> ut.suppress()
    sink()
}
sup.wm <- function(..., mode = "wm") {
    switch(mode,
           w  = suppressWarnings(...),
           m  = suppressMessages(...),
           wm = suppressWarnings(suppressMessages(...)))
}
ut.permargs <- function(...) {
    l.args <- list(...) |> rev()
    Map(function(l, v.ind) l[v.ind],
        l.args,
        mapply(function(v)
               seq_along(v),
               l.args,
               SIMPLIFY = FALSE) |>
        expand.grid())           |>
    rev()
}
ut.save.all <- function() {
    Map(function(n)
        save(list = n,
             file = paste0(n, ".RData"),
```

```
             envir = .GlobalEnv),
        grep("^df|^list",
             ls(".GlobalEnv"), value = TRUE))
}
ut.time <- function() {
    mapply(function(x)
           format(Sys.time(), x),
           c("Date: %Y%m%d %z", "Time: %H%M%S")) |>
    ut.box()
}
ut.pad.dl0 <- function(int, width) {
    sprintf(paste0("%0", width, "d"), int)
}
## parallel +---------------------------------------------+
make.cl <- function(int.cores = parallel::detectCores(),
                    env = environment()) {
    parallel::clusterExport(cl <- parallel::makeCluster(int.cores),
                            ls(envir = env), envir = env)
    cl
}
max.cores <- function() {
    parallel::detectCores()
}
set.env.name <- function() {
    env.parent <- parent.frame(1)
    if(!identical(env.parent, .GlobalEnv)) {
        func.name <- as.character(sys.call(-1))[1]
        if(is(get(func.name, envir = env.parent), "function")) {
            invisible(structure(env.parent,
                                env.name = func.name))
        }
    }
}
## random number generation
gen.rand.int <- function(n) {
    s.int.max <- .Machine[["integer.max"]]
    sample(-s.int.max:s.int.max, n, replace = FALSE)
}
## functional constructs
Compose <- function(...) {
    func.list <- list(...)
    if(!all(unlist(lappy(func.list, is.function))))
        stop("Argument is not a function")
    function(...) {
        Reduce(function(x, f) f(x), func.list, ...)
```

```
    }
}
ComposeM <- function(...) {
    fs <- list(...)
    if(!all(unlist(lappy(fs, is.function))))
        stop("Argument is not a function")
    fs.tail <- fs[-1]
    fs.head <- fs[[1]]
    function(...) {
        Reduce(function(x, f) f(x), fs.tail, fs.head(...))
    }
}
Curry <- function(f, ...) {
    .orig <- list(...)
    function(...) {
        do.call(f, c(.orig, list(...)))
    }
}
CurryL <- function(f, ...) {
    .curried <- as.list(match.call())[c(-1, -2)]
    function(...) {
        .args <- as.list(match.call())[-1]
        eval(substitute(do.call(f, c(.curried, .args))))
    }
}
Filter <- function(f, x, nomatch = NA_integer_) {
    `if`(any((ind <- (unlist(lapply(x, f))) > 0)),
         x[ind], nomatch)
}
Find <- function(f, x, right = FALSE, nomatch = NULL) {
    `if`(any((pos <- Position(f, x, right, nomatch = 0L))),
         x[[pos]], nomatch)
}
Map.TT <- function(f, ...) {
    mapply(FUN = match.fun(f), ...,
           SIMPLIFY  = TRUE,
           USE.NAMES = TRUE)
}
Map.TF <- function(f, ...) {
    mapply(FUN = match.fun(f), ...,
           SIMPLIFY  = TRUE,
           USE.NAMES = FALSE)
}
Map.FT <- function(f, ...) {
    mapply(FUN = match.fun(f), ...,
```

```
            SIMPLIFY  = FALSE,
            USE.NAMES = TRUE)
}
Map.FF <- function(f, ...) {
    mapply(FUN = match.fun(f), ...,
            SIMPLIFY  = FALSE,
            USE.NAMES = FALSE)
}
Map <- Map.FT
Negate.v <- function(...) {
    # Negate a vector.
    !(...)
}
Position <- function(f, x, right = FALSE,
                     nomatch = NA_integer_) {
    f <- match.fun(f)
    ind <- seq_along(x)
    if(right) ind <- rev(ind)
    `for`(i, ind, if(f(x[[i]])) return(i))
    nomatch
}
Swap <- function(f, pos = 2L) {
    f <- match.fun(f)
    pos.coerce <- as.integer(pos) |> suppressWarnings()
    s <- `if`(is.na(pos.coerce),
              {
                  paste0("function(arg1, ...) f(", pos, " = arg1, ...)", collapse = "")
              },
              {
                  if(!all.equal(pos, pos.coerce))
                      stop("pos needs to be either an integer or character string.")
                  paste0("function(",
                         paste0("arg", c(pos, seq(pos.coerce - 1)), collapse = ", "),
                         ", ...)", " f(",
                         paste0("arg", seq(pos), collapse = ", "),
                         ", ...)", collapse = "")
              })
    structure(parse(text = s,
                    keep.source = FALSE)[[1]] |>
              eval(),
              func.name = all.names(sys.call())[-1],
              func = substitute(f))
}
## names
nms <- function(v, v.nm) {
```

```
    # - functional version of assigning names
    # - the main use of this is to facilitate piping in results.
    # v    : object to be renamed
    # v.nm : character vector of new names
    `names<-`(v, v.nm)
}
cnms <- function(v, v.nm) {
    # - functional version of assigning names
    # - the main use of this is to facilitate piping in results.
    # v    : object to be renamed
    # v.nm : character vector of new names
    `colnames<-`(v, v.nm)
}
rnms <- function(v, v.nm) {
    # functional version of assigning row names
    #   - the main use of this is to facilitate piping in results.
    # v    : object to be renamed
    # v.nm : character vector of new names
    `rownames<-`(v, v.nm)
}
nms.sort <- function(obj, ...) {
    # sort an object by its name
    # obj : object to be sorted
    # ... : any arguments to pass to `sort`
    v.nm <- names(obj)
    `if`(is.null(v.nm), obj,
         obj[sort(v.nm, ...)])
}
cnms.sort <- function(obj, ...) {
    # sort an object by its col names
    # obj : object to be sorted
    # ... : any arguments to pass to `sort`
    v.nm <- colnames(obj)
    `if`(is.null(v.nm), obj,
         obj[,sort(v.nm, ...)])
}
rnms.sort <- function(obj, ...) {
    # sort an object by its row names
    # obj : object to be sorted
    # ... : any arguments to pass to `sort`
    v.nm <- rownames(obj)
    `if`(is.null(v.nm), obj,
         obj[sort(v.nm, ...),])
}
## class/type checks
```

```
is.0 <- function(v) {
    # check if object is `0`
    # to return true:
    #   1. the object must not be null
    #   2. have length of 0
    !is.null(v) & (length(v) == 0)
}
is.v <- function(obj, v.class, simplify = any) {
    # multivariate version of is
    # obj:
    #   object to be tested for its class
    # v.class:
    #   character vector of classes
    # simplify:
    #   what function to apply to simplify the result.
    #   default is `any`
    f <- match.fun(simplify)
    f(unlist(mapply(function(s.class)
                    is(obj, s.class),
                    v.class,
                    SIMPLIFY = FALSE,
                    USE.NAMES = TRUE)))
}
class.a <- `class<-`
## comparisons
identical.v <- function(v) {
    if(length(v) == 1)
        return(TRUE)
    bv.na <- is.na(v)
    if(all(bv.na)) {
        warning("Input contains NA.")
        return(TRUE)
    }
    if(any(bv.na)) {
        warning("Input contains NA.")
        return(FALSE)
    }
    all(v[1] == v)
}
all.T <- all
all.F <- function(...) {
    all(!(...))
}
any.T <- any
any.F <- function(...) {
```

```
    any(!(...))
}
ee <- `==`
ne <- `!=`
gt <- `>`
lt <- `<`
gte <- `>=`
lte <- `<=`

## flow control
if.t <- function(b.test, v.true, v.false) {
    # `if` that switches to binary or ternary version.
    # - if `v.false` is missing, switches to binary version.
    # - ternary version if like ifelse but accepts only singular boolean.

    # notes:
    # - the primary intended use of this function is to facilitate the use of if in pipes.
    # - together with the `Swap` function, allows flexible piping of arguments.
    if(missing(v.false)) {
        if(b.test) {
            v.true
        }
    } else {
        if(b.test) {
            v.true
        } else {
            v.false
        }
    }
}
sw.b <- function(...) {
    b <- c(...)
    s.len <- length(b)
    if(any(!is.logical(b), (s.len == 0), (sum(b) > 1)))
        stop("Error")
    switch(which(c((s.len == 1),
                   (s.len > 1))),
           ifelse(b, 1, 2),
           which(b))
}
## subsetting
ss.b <- function(v, i) {
    `[`(v, i)
}
ss.bb <- function(v, i) {
```

```
    `[[`(v, i)
}
ss.e.nm <- function(obj, v.exclude.names) {
    # subset, excluding by names
    v.obj.names <- names(obj)
    if(is.null(v.obj.names)) {
        stop("Input object has no names.")
    } else {
        b.include <- match(v.obj.names,
                           v.exclude.names,
                           nomatch = 0) == 0
        b.exclude <- !b.include
        if(all(b.exclude))
            stop("All items excluded.")
        if(any(b.include))
            obj[b.include]
    }
}
ssa.b <- function(v, i, v.rep) {
    `[<-`(v, i, v.rep)
}
ssa.bb <- function(v, i, v.rep) {
    `[[<-`(v, i, v.rep)
}
## attributes
attr.exa <- function(obj) {
    attributes(obj)
}
attr.ex1 <- function(obj, attr.nm) {
    attr.exa(obj)[[attr.nm]]
}
attr.set <- function(obj, l.attr) {
    do.call(structure, c(obj, l.attr))
}
attr2env <- function(obj, env = parent.frame()) {
    l <- attributes(obj)
    b.include <- match(names(l), c("dim", "names"), nomatch = 0) == 0
    if(any(b.include))
        list2env(l[b.include], envir = env)
}
set.list.index <- function(l) {
    # create index
    l.index <- list(i = seq_along(l))
    v.names <- names(l)
    if(!is.null(v.names))
```

```
        l.index <- c(l.index, list(names = v.names))

    # index is now a list of lists of the indexes.
    # - this is a linked list and will need to be restructured
    # - restructure to be 1 list for each item present in l containing the attributes.
    l.index.rstruc <- do.call(Map,
                              c(list(list),
                                Map(as.list, l.index)))

    # attach the attributes
    mapply(function(i, l.attr)
           structure(i, index = l.attr),
           l, l.index.rstruc,
           SIMPLIFY = FALSE,
           USE.NAMES = TRUE)
}
## glm
get.glm.dfres <- function(mo) {
    summary(mo)[["df"]][2]
}
get.glm.np <- function(mo) {
    v.df <- summary(mo)[["df"]]
    list(n = sum(v.df[2:3]), p = v.df[3])
}
get.glm.var <- function(mo) {
    v <- formula(mo) |> all.vars()
    list(dv = v[1], iv = v[-1])
}
# packages
library(compiler)
# options
enableJIT(3)
set.seed(8)
options(list(stringsAsFactors = FALSE,
             save.defaults = list(compress = "xz",
                                  compression_level = 9),
             contrasts = c(unordered = "contr.treatment",
                           ordered = "contr.treatment"),
             max.print = 40))
# save functions
list.ut <- ut.get.all()

# clear workspace
ut.clear.globalwspc()
# turn off display devices
```

```
    ut.clear.dispdev()
    # import functions
    list2env(list.ut, envir = .GlobalEnv)

    # clear environment and show time
    ut.clear.disp()
    ut.time()

    # return functions
    list.ut
}
```

G.2 Model- and simulation-based estimator

The *gen.funcs.seqmed* function generates the functions for the main programme. Use the command *list2env(gen.funcs.seqmed())* to insert these functions into the present global environment. The main workhorse of the programme consists of two functions: *mo.med* and *sim.med* for model- and simulation- based estimators respectively. Sensitivity analysis is done using the *sim.med.sa* function.

The data needs to be prepared with the following column order: outcome of interest, treatment indicator, any covariates, second mediator, first mediator. The models for the outcome, second mediator and first mediator then need to be fitted and the resulting models then need to be put into a list with the aforementioned order. This list can then be used in the functions for estimation.

The arguments for each of the function for estimation are as follows:

```
Using the notation of:
    - y for outcomes
    - r for treatment indicator
    - cv for covariates
    - m1 for first mediator
    - m2 for second mediator

Before using the functions, the y, m2 and m1 models need to be first fitted with the glm command.
The data used to fit the models must have the following variable order:
```

```
    y, r, cv, m2, m1
The fitted models should then be put into a list of models with the following order:
    y, m2, m1
The list of models are referred to as l.mo.
The values shown below beside the functions are the default values.

# mo.med
mo.med(l.mo,
       s.ci       = 0.95,
       int.boot   = 1e4L,
       s.seed     = gen.seeds(1),
       b.parallel = "auto",
       int.cores  = "max")

l.mo is the list of fitted models.
int.boot is the number of bootstraps to use for the confidence interval.
s.seed is the starting seed for the simulation.
b.parallel is the switch to enable or disable parallel processing.
    - "auto" will check the number of simulations requested and
       if it is above 10000, turn on parallel processing.
int.cores is the number of cores to use if parallel processing is turned on.

# sim.med and sim.med.sa
sim.med(l.mo,
        s.cof.mth = "mvn",
        s.ci = 0.95,
        int.sims = 1e4L,
        s.seed = gen.seeds(1),
        b.parallel = "auto",
        int.cores = "max",
        b.raw = TRUE)
sim.med.sa(l.mo,
           s.cof.mth = "mvn",
           s.ci = 0.95,
           int.sims = 1e1L,
           s.seed = gen.seeds(1),
           b.parallel = "auto",
           int.cores = "max")

l.mo is the list of fitted models.
s.cof.mth can have the values:
    - "mvn" to use the multivariate normal method for simulation.
    - "bootstrap" to use the bootstrap method for simulation.
int.sims is the number of simulations.
s.seed is the starting seed for the simulation.
```

```
b.parallel is the switch to enable or disable parallel processing.
    - "auto" will check the number of simulations requested and
       if it is above 10000, turn on parallel processing.
int.cores is the number of cores to use if parallel processing is turned on.
b.raw allows the raw data of the simulations to be exported.


gen.funcs.seqmed <- function() {
    ## utilities
    ut.box <- function(v.contents, chr.head = "+", chr.sep = "-",
                       chr.tail = "+", chr.wall = "|", s.align = "l") {
        if(class(v.contents) |>
           match(c("integer", "numeric", "character"),
                 nomatch = 0) |>
           (any |> Negate())())
            stop("Incorrect contents format.")

        s.contents.lenmax <- nchar(v.contents) |> max()
        s.box.lenmax <- length(c(chr.head, chr.sep, chr.tail)) |> sum()
        if(s.contents.lenmax < s.box.lenmax)
            s.contents.lenmax <- s.box.lenmax
        s.contents.lenpadded <- s.contents.lenmax + 2
        s.contents.lenpadded <- (4 - (s.contents.lenpadded %% 4)) + s.contents.lenpadded

        s.contents.lenpadded.use <-
            switch(s.align,
                   l = -s.contents.lenpadded,
                   r = s.contents.lenpadded,
                   c = s.contents.lenpadded/2)
        v.messages <-
            switch((s.align == "c") + 1,
                   sprintf(paste0(chr.wall, " %*s ", chr.wall),
                           s.contents.lenpadded.use, v.contents),
                   {
                       l.contents <-
                           Map(function(x)
                               {
                                   s.len <- length(x) %/% 2
                                   Map(function(v) paste0(v, collapse = ""),
                                       list(head(x, s.len), tail(x, -s.len)))
                               }, strsplit(v.contents, ""))

                       mapply(function(x)
                              {
                                  paste0(sprintf(paste0(chr.wall, " %*s"),
                                                 s.contents.lenpadded.use, x[[1]]),
```

```
                                    sprintf(paste0("%*s ", chr.wall),
                                            -s.contents.lenpadded.use, x[[2]]),
                                    collapse = "")
                            }, l.contents)
                })
    v.sep <- paste0(c(chr.head,
                      rep(chr.sep,
                          s.contents.lenpadded +
                              ((nchar(c(" ", chr.wall)) * 2) |> sum()) -
                              (nchar(c(chr.head, chr.tail))  |> sum())),
                      chr.tail),
                    collapse = "")
    mapply(message, c(v.sep, v.messages, v.sep)) |> invisible()
}
ut.clear.disp <- function() {
    if(!is.null(dev.list())) {
        invisible(mapply(dev.off, dev.list()))
    }
    switch(commandArgs()[1],
           RStudio = cat("\014"),
           system("clear"))
}
ut.clear.dispdev <- function() {
if(!is.null(dev.list()))
    ut.silent(mapply(dev.off, dev.list()))
}
ut.clear.globalwspc <- function() {
    local(rm(list = ls()), envir = .GlobalEnv)
}
ut.get.all <- function() {
    env <- parent.frame(1)
    mget(ls(envir = env), envir = env)
}
ut.suppress <- function(...) {
    suppressMessages(...) |> suppressWarnings()
}
ut.silent <- function(...) {
    sink(nullfile())
    try(..., silent = TRUE, outFile = nullfile()) |> ut.suppress()
    sink()
}
sup.wm <- function(..., mode = "wm") {
    switch(mode,
           w  = suppressWarnings(...),
           m  = suppressMessages(...),
```

```
            wm = suppressWarnings(suppressMessages(...)))
}
ut.permargs <- function(...) {
    l.args <- list(...) |> rev()
    Map(function(l, v.ind) l[v.ind],
        l.args,
        mapply(function(v)
               seq_along(v),
               l.args,
               SIMPLIFY = FALSE) |>
        expand.grid())           |>
    rev()
}
ut.save.all <- function() {
    Map(function(n)
        save(list = n,
             file = paste0(n, ".RData"),
             envir = .GlobalEnv),
        grep("^df|^list",
             ls(".GlobalEnv"), value = TRUE))
}
ut.time <- function() {
    mapply(function(x)
           format(Sys.time(), x),
           c("Date: %Y%m%d %z", "Time: %H%M%S")) |>
    ut.box()
}
ut.pad.dl0 <- function(int, width) {
    sprintf(paste0("%0", width, "d"), int)
}
## parallel +-------------------------------------------+
make.cl <- function(int.cores = parallel::detectCores(),
                    env = environment()) {
    parallel::clusterExport(cl <- parallel::makeCluster(int.cores),
                            ls(envir = env), envir = env)
    cl
}
max.cores <- function() {
    parallel::detectCores()
}
set.env.name <- function() {
    env.parent <- parent.frame(1)
    if(!identical(env.parent, .GlobalEnv)) {
        func.name <- as.character(sys.call(-1))[1]
        if(is(get(func.name, envir = env.parent), "function")) {
```

```
            invisible(structure(env.parent,
                                env.name = func.name))
        }
    }
}
## random number generation
gen.rand.int <- function(n) {
    s.int.max <- .Machine[["integer.max"]]
    sample(-s.int.max:s.int.max, n, replace = FALSE)
}
## functional constructs
Compose <- function(...) {
    func.list <- list(...)
    if(!all(unlist(lappy(func.list, is.function))))
        stop("Argument is not a function")
    function(...) {
        Reduce(function(x, f) f(x), func.list, ...)
    }
}
ComposeM <- function(...) {
    fs <- list(...)
    if(!all(unlist(lappy(fs, is.function))))
        stop("Argument is not a function")
    fs.tail <- fs[-1]
    fs.head <- fs[[1]]
    function(...) {
        Reduce(function(x, f) f(x), fs.tail, fs.head(...))
    }
}
Curry <- function(f, ...) {
    .orig <- list(...)
    function(...) {
        do.call(f, c(.orig, list(...)))
    }
}
CurryL <- function(f, ...) {
    .curried <- as.list(match.call())[c(-1, -2)]
    function(...) {
        .args <- as.list(match.call())[-1]
        eval(substitute(do.call(f, c(.curried, .args))))
    }
}
Filter <- function(f, x, nomatch = NA_integer_) {
    `if`(any((ind <- (unlist(lapply(x, f))) > 0)),
         x[ind], nomatch)
```

```
}
Find <- function(f, x, right = FALSE, nomatch = NULL) {
    `if`(any((pos <- Position(f, x, right, nomatch = 0L))),
         x[[pos]], nomatch)
}
Map.TT <- function(f, ...) {
    mapply(FUN = match.fun(f), ...,
           SIMPLIFY  = TRUE,
           USE.NAMES = TRUE)
}
Map.TF <- function(f, ...) {
    mapply(FUN = match.fun(f), ...,
           SIMPLIFY  = TRUE,
           USE.NAMES = FALSE)
}
Map.FT <- function(f, ...) {
    mapply(FUN = match.fun(f), ...,
           SIMPLIFY  = FALSE,
           USE.NAMES = TRUE)
}
Map.FF <- function(f, ...) {
    mapply(FUN = match.fun(f), ...,
           SIMPLIFY  = FALSE,
           USE.NAMES = FALSE)
}
Map <- Map.FT
Negate.v <- function(...) {
    # Negate a vector.
    !(...)
}
Position <- function(f, x, right = FALSE,
                     nomatch = NA_integer_) {
    f <- match.fun(f)
    ind <- seq_along(x)
    if(right) ind <- rev(ind)
    `for`(i, ind, if(f(x[[i]])) return(i))
    nomatch
}
Swap <- function(f, pos = 2L) {
    f <- match.fun(f)
    pos.coerce <- as.integer(pos) |> suppressWarnings()
    s <- `if`(is.na(pos.coerce),
              {
                  paste0("function(arg1, ...) f(", pos, " = arg1, ...)", collapse = "")
              },
```

```
            {
                if(!all.equal(pos, pos.coerce))
                    stop("pos needs to be either an integer or character string.")
                paste0("function(",
                       paste0("arg", c(pos, seq(pos.coerce - 1)), collapse = ", "),
                       ", ...)", " f(",
                       paste0("arg", seq(pos), collapse = ", "),
                       ", ...)", collapse = "")
            })
    structure(parse(text = s,
                    keep.source = FALSE)[[1]] |>
            eval(),
            func.name = all.names(sys.call())[-1],
            func = substitute(f))
}
## names
nms <- function(v, v.nm) {
    # - functional version of assigning names
    # - the main use of this is to facilitate piping in results.
    # v    : object to be renamed
    # v.nm : character vector of new names
    `names<-`(v, v.nm)
}
cnms <- function(v, v.nm) {
    # - functional version of assigning names
    # - the main use of this is to facilitate piping in results.
    # v    : object to be renamed
    # v.nm : character vector of new names
    `colnames<-`(v, v.nm)
}
rnms <- function(v, v.nm) {
    # functional version of assigning row names
    #   - the main use of this is to facilitate piping in results.
    # v    : object to be renamed
    # v.nm : character vector of new names
    `rownames<-`(v, v.nm)
}
nms.sort <- function(obj, ...) {
    # sort an object by its name
    # obj : object to be sorted
    # ... : any arguments to pass to `sort`
    v.nm <- names(obj)
    `if`(is.null(v.nm), obj,
         obj[sort(v.nm, ...)])
}
```

```
cnms.sort <- function(obj, ...) {
    # sort an object by its col names
    # obj : object to be sorted
    # ... : any arguments to pass to `sort`
    v.nm <- colnames(obj)
    `if`(is.null(v.nm), obj,
         obj[,sort(v.nm, ...)])
}
rnms.sort <- function(obj, ...) {
    # sort an object by its row names
    # obj : object to be sorted
    # ... : any arguments to pass to `sort`
    v.nm <- rownames(obj)
    `if`(is.null(v.nm), obj,
         obj[sort(v.nm, ...),])
}
## class/type checks
is.0 <- function(v) {
    # check if object is `0`
    # to return true:
    #   1. the object must not be null
    #   2. have length of 0
    !is.null(v) & (length(v) == 0)
}
is.v <- function(obj, v.class, simplify = any) {
    # multivariate version of is
    # obj:
    #   object to be tested for its class
    # v.class:
    #   character vector of classes
    # simplify:
    #   what function to apply to simplify the result.
    #   default is `any`
    f <- match.fun(simplify)
    f(unlist(mapply(function(s.class)
                    is(obj, s.class),
                    v.class,
                    SIMPLIFY = FALSE,
                    USE.NAMES = TRUE)))
}
class.a <- `class<-`
## comparisons
identical.v <- function(v) {
    if(length(v) == 1)
        return(TRUE)
```

```
    bv.na <- is.na(v)
    if(all(bv.na)) {
        warning("Input contains NA.")
        return(TRUE)
    }
    if(any(bv.na)) {
        warning("Input contains NA.")
        return(FALSE)
    }
    all(v[1] == v)
}
all.T <- all
all.F <- function(...) {
    all(!(...))
}
any.T <- any
any.F <- function(...) {
    any(!(...))
}
ee <- `==`
ne <- `!=`
gt <- `>`
lt <- `<`
gte <- `>=`
lte <- `<=`

## flow control
if.t <- function(b.test, v.true, v.false) {
    # `if` that switches to binary or ternary version.
    # - if `v.false` is missing, switches to binary version.
    # - ternary version if like ifelse but accepts only singular boolean.

    # notes:
    # - the primary intended use of this function is to facilitate the use of if in pipes.
    # - together with the `Swap` function, allows flexible piping of arguments.
    if(missing(v.false)) {
        if(b.test) {
            v.true
        }
    } else {
        if(b.test) {
            v.true
        } else {
            v.false
        }
```

```
    }
}
sw.b <- function(...) {
    b <- c(...)
    s.len <- length(b)
    if(any(!is.logical(b), (s.len == 0), (sum(b) > 1)))
        stop("Error")
    switch(which(c((s.len == 1),
                   (s.len > 1))),
           ifelse(b, 1, 2),
           which(b))
}
## subsetting
ss.b <- function(v, i) {
    `[`(v, i)
}
ss.bb <- function(v, i) {
    `[[`(v, i)
}
ss.e.nm <- function(obj, v.exclude.names) {
    # subset, excluding by names
    v.obj.names <- names(obj)
    if(is.null(v.obj.names)) {
        stop("Input object has no names.")
    } else {
        b.include <- match(v.obj.names,
                           v.exclude.names,
                           nomatch = 0) == 0
        b.exclude <- !b.include
        if(all(b.exclude))
            stop("All items excluded.")
        if(any(b.include))
            obj[b.include]
    }
}
ssa.b <- function(v, i, v.rep) {
    `[<-`(v, i, v.rep)
}
ssa.bb <- function(v, i, v.rep) {
    `[[<-`(v, i, v.rep)
}
## attributes
attr.exa <- function(obj) {
    attributes(obj)
}
```

```
attr.ex1 <- function(obj, attr.nm) {
    attr.exa(obj)[[attr.nm]]
}
attr.set <- function(obj, l.attr) {
    do.call(structure, c(obj, l.attr))
}
attr2env <- function(obj, env = parent.frame()) {
    l <- attributes(obj)
    b.include <- match(names(l), c("dim", "names"), nomatch = 0) == 0
    if(any(b.include))
        list2env(l[b.include], envir = env)
}
set.list.index <- function(l) {
    # create index
    l.index <- list(i = seq_along(l))
    v.names <- names(l)
    if(!is.null(v.names))
        l.index <- c(l.index, list(names = v.names))

    # index is now a list of lists of the indexes.
    # - this is a linked list and will need to be restructured
    # - restructure to be 1 list for each item present in l containing the attributes.
    l.index.rstruc <- do.call(Map,
                              c(list(list),
                                Map(as.list, l.index)))

    # attach the attributes
    mapply(function(i, l.attr)
           structure(i, index = l.attr),
           l, l.index.rstruc,
           SIMPLIFY = FALSE,
           USE.NAMES = TRUE)
}
## glm
get.glm.dfres <- function(mo) {
    summary(mo)[["df"]][2]
}
get.glm.np <- function(mo) {
    v.df <- summary(mo)[["df"]]
    list(n = sum(v.df[2:3]), p = v.df[3])
}
get.glm.var <- function(mo) {
    v <- formula(mo) |> all.vars()
    list(dv = v[1], iv = v[-1])
}
```

```
# helper +------------------------------------------------------------+
## environment and variables management +------------------+
rm.attr <- function(o) {
    attributes(o) <- attributes(o)[c("dim", "dimnames")]
    o
}
add.attr <- function(o, l) {
    `attributes<-`(o, c(l, attributes(o)))
}
add.nm <- function(v, v.nm) {
    if.t(any(!is.vector(v),
             !is.character(v.nm),
             !identical(length(v), length(v.nm))),
         stop("Error"),
       nms(v, v.nm))
}
add.cnm <- function(d, v.nm) {
    if(any(!any(class(d) %in% c("matrix", "data.frame")), !is.character(v.nm)))
        stop("Error")
    if(!identical(ncol(d), length(v.nm)))
        stop("Error")
    cnms(d, v.nm)
}
add.rnm <- function(d, v.nm) {
    if(any(!any(class(d) %in% c("matrix", "data.frame")), !is.character(v.nm)))
        stop("Error")
    if(!identical(nrow(d), length(v.nm)))
        stop("Error")
    rnms(d, v.nm)
}
list.var <- function(l) {
    if.t(class(l) != "list",
         as.list(l), l)               |>
    list2env(envir = parent.frame(1)) |>
    ut.silent()
}
class.check <- function(obj, v.valid.classes, invert = FALSE) {
    v.checks <- match(obj, v.valid.classes, nomatch = 0)
    if(invert) {
        all(v.checks == 0)
    } else {
        any(v.checks > 0)
    }
}
```

```
setup.env <- function(list.models,
                      int.iter,
                      b.parallel = "auto",
                      int.cores = "max",
                      canonical = FALSE)
{
    # hardcoded values
    int.parallel.threshold <- 1e4L

    # function to setup consistent environment across functions
    # argumemts:
    #     list.models : list of fitted models
    # each fitted model must have as its first term the treatment variable.

    # get all the terms used in all the models
    l.nm.av <- Map(function(mo) formula(mo) |> all.vars(), list.models)
    # get dependent variables
    v.nm.dv <- Map(ss.b, l.nm.av, 1) |> unlist()

    # extract degrees of freedom
    m.df <- do.call(rbind, Map(ss.bb, Map(summary, list.models), "df"))

    # get the number of coefficients
    # - order them from most tp least number
    # - this order should correspond to:
    #   + y model
    #   + m2 model
    #   + m1 model
    p <- m.df[,3] |> nms(v.nm.dv) |> sort(decreasing = TRUE)

    # get the sample size from degrees of freedom and
    # check that the sample size is the same across models.
    v.n <- (m.df[,2] + m.df[,3]) |> unique()
    n <- if.t(length(v.n) > 1,
              stop("Unequal sample sizes across models."),
              v.n)

    # get the residual degrees of freedom
    df.res <- (m.df[,2] |> nms(v.nm.dv))[names(p)]

    l.np <- list(n = n, p = p, df.res = df.res)

    # name the models and arrange them in descending order of number of coefficients
    list.models <- nms(list.models, v.nm.dv)[names(p)]
```

```
# dataset
d <- model.frame(list.models[[1]])

# replace dv with the names of p as the right order of dv
v.nm.dv <- names(p)
v.nm.m <- v.nm.dv[-1]

# extract iv and identify common terms
l.nm.iv <- Map(ss.b, l.nm.av, -1)
v.nm.common <- Reduce(intersect, l.nm.iv)

# extract treatment variable
## the first term should be the treatment variable.
v.nm.r <- Map(ss.b, l.nm.iv, 1) |> unlist()
## making sure that the first term of each model is the same.
if(!identical.v(v.nm.r))
    stop("Error with treatment variable in models.")
s.nm.r <- v.nm.r[1] |> unname()
## get unique values of treatment
v.r <- d[[s.nm.r]] |>
       unique()    |>
       sort(decreasing = TRUE)

# get covariates
b.cv <- match(v.nm.common, s.nm.r, nomatch = 0) == 0
v.nm.cv <- if.t(any(b.cv), v.nm.common[b.cv], NA)

if(canonical) {
    # giving canonical names to dv, r, iv
    names(v.nm.dv) <- c("y", sprintf("m%02d", seq_along(v.nm.dv[-1])))
    names(s.nm.r) <- "r"
    names(v.nm.cv) <- sprintf("cv%02d", seq_along(v.nm.cv))
} else {
    names(v.nm.dv) <- v.nm.dv
    names(s.nm.r) <- s.nm.r
    names(v.nm.cv) <- v.nm.cv
}

# setup new iv and allv names
v.nm.iv <- c(s.nm.r, v.nm.cv)
v.nm.av <- c(v.nm.dv, v.nm.iv)

# identify family of each model
v.fam <- Map(ss.bb,
             Map(family, list.models),
```

```
            "family") |>
         unlist()      |>
         Swap(gsub, 3)("^(.).*", "\\1")
s.fam <- paste0(v.fam, collapse = "")
## use abbreviations to identify models:
##     g for gaussian
##     b for binary
l.fam <- local({
    v.abv <- c("g", "b")
    Map(function(x)
        structure(v.abv == x,
                  names = v.abv),
        v.fam)})

# number of mediators and models
n.mo <- length(list.models)
n.med <- n.mo - 1

# process parallel parameters
if(!is.logical(b.parallel)) {
    if(!(b.parallel == "auto"))
        stop("Invalid parallel option")
    b.parallel <- int.iter >= int.parallel.threshold
}
int.cores <-
    if.t(b.parallel,
         if.t(int.cores == "max",
              parallel::detectCores(),
              {
                  cores <- as.integer(int.cores)
                  if.t(is.na(cores),
                       {
                           warning("Invalid cores option, setting it to maximum number of cores.")
                           parallel::detectCores()
                       }, cores)
              }), 1L)
# gather results
l.base <- list(list.models,
               d          = d,
               n.mo       = n.mo,
               n.med      = n.med,
               s.n        = n,
               p          = p,
               df.res     = df.res,
               s.fam      = s.fam,
```

```
                    l.fam      = l.fam,
                    v.nm.dv    = v.nm.dv,
                    v.nm.m     = v.nm.m,
                    v.nm.iv    = v.nm.iv,
                    s.nm.r     = s.nm.r,
                    v.r        = v.r,
                    v.nm.cv    = v.nm.cv,
                    v.nm.av    = v.nm.av,
                    l.nm.av    = l.nm.av,
                    int.iter   = int.iter,
                    b.parallel = b.parallel,
                    int.cores  = int.cores)

    l.base[[1]] <- do.call(structure, l.base)
    names(l.base)[1] <- "l.mo"

    list2env(c(list(l.base = l.base), l.base),
             envir = parent.frame(1)) |>
    ut.silent()
}
## objects manipulation +---------------------------------+
mat.dupe <- function(m, fill = NA) {
    m[] <- fill
    class(m) <- class(fill)
    m
}
mat.mirr <- function(m, s.orient) {
    m <- cbind(m)
    v.dim <- dim(m)
    cbind(switch(s.orient,
                 lr = m[,v.dim[2]:1],
                 ud = m[v.dim[1]:1,]))
}
mat.t.list <- function(m, by, ret) {
    if(!any(by %in% c(1, 2))) stop("Error")
    l <- unlist(apply(m, by, list), recursive = FALSE)
    switch(ret, v = l, m = Map(rbind, l))
}
vec.t.list <- function(v, v.len) {
    if(length(v) != sum(v.len)) stop("Error")
    env <- environment()
    Map(function(s)
        {
            v.ind <- seq(s)
            v.ret <- v[v.ind]
```

```
            assign("v", v[-v.ind], envir = env)
            v.ret
        }, v.len)
}
vec.t.matc <- function(v) {
    v |> as.matrix()
}
vec.t.matr <- function(v) {
    v |> as.matrix() |> t()
}
combi <- function(vm1, vm2) {
    m1 <- cbind(vm1)
    m2 <- cbind(vm2)
    n.m1 <- nrow(m1)
    n.m2 <- nrow(m2)
    m <- cbind(m1[sort(rep(1:n.m1, n.m2)),],
               m2[rep(1:n.m2, n.m1),])
    m[order(apply(m, 1, paste, collapse = "")),]
}
combi.bin <- function(v.nm, s.ret) {
    l.bin.resp <- rep(list(0:1), length(v.nm))
    m.combi <- add.cnm(cbind(Reduce(combi, l.bin.resp)), v.nm)
    l.combi <- mat.t.list(m.combi, 1, s.ret)
}
str.half <- function(v, s.type = "v", b.clp = TRUE) {
    s.len <- length(v)
    if((s.len %% 2) != 0) stop("Error")
    s.lenh <- s.len/2
    l <- Map(function(v.ind) v[v.ind], list(1:s.lenh, (s.lenh + 1):s.len))
    if(b.clp) {
        l <- Map(paste0, l, collapse = "")
    }
    if(s.type == "v") {
        l <- do.call(c, l)
    }
    l
}
str.pad0 <- function(v, s.len = max(nchar(v))) {
    sprintf(paste0("%0", s.len, "d"), v)
}
mo.upd <- function(mo, fml) {
    update(mo,
           formula = fml,
           family = family(mo),
           data = model.frame(mo))
```

```
}
## comparisons and flow control +-------------------------+
pblapply.sw <- function(b.parallel, int.cores, env, ...) {
    f <- pbapply::pblapply
    args <- list(...)
    if(b.parallel) {
        cl <- make.cl(int.cores = int.cores, env = environment())
        args <- c(args, list(cl = cl))
    }
    obj <- do.call(f, args)
    if(b.parallel) parallel::stopCluster(cl)
    obj
}
pbMap <- local({
    f <- pbapply::pbmapply
    formals(f)[["SIMPLIFY"]] <- FALSE
    f
})
# set up environment +-----------------------------------------------+
## glm families +---------------------------------------+
get.fam.raw <- function(mo) {
    with(family(mo), c(family, link))
}
get.fam.bool <- function(mo) {
    b <- get.fam.raw(mo)[1] == "gaussian"
    structure(c(b, !b), .Names = c("g", "b"))
}
get.fam.l <- function(l.mo) {
    attr2env(l.mo)
    add.nm(Map(get.fam.bool, l.mo), v.nm.dv)
}
get.fam.v <- function(l.mo) {
    b <- unlist(unname(get.fam.l(l.mo)))
    paste0(names(b)[b], collapse = "")
}
get.mo.dep <- function(l.mo) {
    attr2env(l.mo)
    l.dep <- Map(function(v.nm, v.var)
                 {
                     v <- intersect(v.nm.dv, v.var[-(1:2)])
                     if(is.0(v)) {
                         NULL
                     } else {
                         v
                     }
```

```
                }, v.nm.dv, l.nm.av)
    l.dep.n <- mapply(function(v) length(v), l.dep)
    list(dep = l.dep, n = l.dep.n, n.sorted = sort(l.dep.n))
}
set.exv <- function(v.nm.dv, v.ev, l.fam) {
    mapply(function(s.nm, s.ev, vb.fam)
           {
               b <- if(s.ev == "auto") {
                   if(s.nm == "y") {
                       TRUE
                   } else {
                       vb.fam[2]
                   }
               } else {
                   s.ev
               }
               structure(b, .Names = "")
           }, v.nm.dv, v.ev, l.fam)
}
## get n (sample size) & p (variables) +------------------+
get.np <- function(l.mo, s.type) {
    m.np <- mapply(function(mo) dim(model.matrix(mo)), l.mo)
    if(!identical.v(m.np[1,]))
        stop("Error: Unequal sample sizes across models.")
    switch(s.type,
           n = m.np[1,1],
           np =
           {
               v.nm <- mapply(function(mo)
                              as.character(formula(mo))[2], l.mo)
               switch(s.type,
                      p = structure(m.np[2,], .Names = v.nm),
                      np = list(n = m.np[1,1],
                                p = structure(m.np[2,], .Names = v.nm)))
           })
}
## confidence intervals +---------------------------------+
gen.ci <- function(s.ci) {
    if(any(!is.numeric(s.ci),
           length(s.ci) > 1,
           s.ci < 0,
           s.ci > 1)) stop("Error")
    s.lb <- (1 - s.ci)/2
    s.ub <- s.ci + s.lb
    c(lb = s.lb, ub = s.ub)
```

```
}
ci.perc <- function(v, s.ci) {
    quantile(v, probs = gen.ci(s.ci), type = 7)
}

# mathematical +-------------------------------------------------+
dif <- function(...) Reduce(`-`, ...)
div <- function(...) Reduce(`/`, ...)
cumdif <- function(...) Reduce(`-`, ..., accumulate <- TRUE)
cumdiv <- function(...) Reduce(`/`, ..., accumulate <- TRUE)
gmean <- function(v) v |> log() |> mean() |> exp()
logit <- qlogis
expit <- plogis
gen.seeds <- function(int.sims, s.seed = NULL) {
    # generate seeds
    s.int.max <- .Machine[["integer.max"]]
    set.seed(s.seed)
    sample((-s.int.max):s.int.max, int.sims,
           replace = FALSE)
}

# model fitting and parmeters extraction +--------------------------+
chr.ap.nm <- function(s.nm, v) {
    paste0(s.nm, "~", v)
}
chr.clp <- function(v, type = "") {
    # string utility: collapse
    paste0(v, collapse = type)
}
chr.pad <- function(v, s.chr = " ") {
    # string utility: pad
    v.nchar <- nchar(v)
    v.padwd <- max(v.nchar) - v.nchar
    mapply(function(s, s.wd) paste0(c(s, rep(s.chr, s.wd)), collapse = "") , v, v.padwd)
}
chr.rp.i <- function(v) {
    # Replace "(Intercept)" with 1
    gsub("^\\(Intercept\\)$", "1", v)
}
chr.sur <- function(s, s.chr = "\"") {
    # string utility: surround
    chr.clp(c(s.chr, s, s.chr))
}
ind <- function(v.char) {
    v.uniq <- unique(v.char)
```

```
    structure(1:length(v.uniq), .Names = v.uniq)
}
m.flat <- function(m, na.rm = FALSE) {
    v <- dim(m)
    v.r <- rep(1:v[1], v[2])
    v.c <- sort(rep(1:v[2], v[1]))
    m.mask <- cbind(v.r, v.c)

    v.nm.r <- rownames(m)
    v.nm.c <- colnames(m)

    if(!is.null(v.nm.r)) v.r <- v.nm.r[v.r]
    if(!is.null(v.nm.c)) v.c <- v.nm.c[v.c]
    d <- data.frame(row = v.r, col = v.c, val = m[m.mask])
    if(na.rm) {
        d <- na.omit(d)
    }
    d[with(d, order(col, row)),]
}
m.rcon <- function(d) {
    v.nm.r <- ind(d[,"row"])
    v.nm.c <- ind(d[,"col"])

    m <- matrix(0,
                nrow = length(v.nm.r),
                ncol = length(v.nm.c),
                dimnames = list(names(v.nm.r),
                                names(v.nm.c)))
    m.mask <- cbind(row = v.nm.r[d[,"row"]], col = v.nm.c[d[,"col"]])
    m[m.mask] <- d[,"val"]

    b.nm.r.int <- !suppressWarnings(any(is.na(as.integer(names(v.nm.r)))))
    b.nm.c.int <- !suppressWarnings(any(is.na(as.integer(names(v.nm.c)))))

    if(b.nm.r.int) {
        rownames(m) <- NULL
    } else {
        m <- m[sort(rownames(m)),]
    }
    if(b.nm.c.int) {
        colnames(m) <- NULL
    } else {
        m <- m[,sort(colnames(m))]
    }
    m
```

```
}
m.symm <- function(m) {
    m[upper.tri(m)] <- t(m)[upper.tri(m)]
    m
}
rm.null.1 <- function(i) {
    i[!mapply(function(l1) is.null(l1), i)]
}
rm.null <- function(l) {
    # list utility: remove null elements
    Map(function(l1)
        {
            if(is.list(l1)) {
                rm.null.1(l1)
            } else {
                l1
            }
        }, rm.null.1(l))
}
catf <- function(v, lineend, file) {
    writeBin(paste0(c(v, ""), collapse = lineend) |>
             charToRaw(), file)
}
## extract parameters +----------------------------------+
mp.ext.t0.cf <- function(mp.out, var.est = "mplus") {
    mp.varindex <- attributes(mp.out)[["varindex"]]
    l.mo <- mp.out[["models"]]
    attr2env(l.mo)
    v.nm.u <- mp.out[["raw"]][["nm.u"]]
    v.nm.n <- mp.out[["raw"]][["nm.n"]]
    v.nm.o <- c(names(v.nm.n), v.nm.u, 1)
    names(v.nm.o) <- toupper(c(v.nm.n, v.nm.u, 1))
    d <- within(mp.out[["est"]][["parameters"]][["unstandardized"]],
                est[paramHeader == "Thresholds"] <-
                    est[paramHeader == "Thresholds"] * (-1))
    v.parmheaders <-
        c("Means",
          "Intercepts",
          "Thresholds",
          "Variances",
          "Residual\\.Variances")
    v.swap.reg <- paste0(c("\\.BY$",
                           paste0("^",
                                  c("", v.parmheaders),
                                  "$")),
```

```
                    collapse = "|")
v.headers <-
    mapply(function(d.i)
           {
               d.new <- d.i
               if(grepl(v.swap.reg,
                        d.i[,"paramHeader"])) {
                   d.new[,"paramHeader"] <- d.i[,"param"]
                   d.new[,"param"] <- d.i[,"paramHeader"]
               }
               d.new <- within(d.new,
                               {
                                   paramHeader <- gsub("\\$1$", "", paramHeader)
                                   paramHeader <- gsub("\\.ON$", "", paramHeader)
                                   paramHeader <- gsub("\\.WITH$", "", paramHeader)
                                   param <- gsub("\\.BY$", "", param)
                                   param <- gsub("^Means$|^Intercepts$|^Thresholds$", "1", param)
                                   b.var <- grepl("Variances$", param)
                                   param[b.var] <- paramHeader[b.var]
                                   paramHeader <- v.nm.o[paramHeader]
                                   param <- v.nm.o[param]
                                   rm(b.var)
                               })
                       paste0(d.new, collapse = "~")
                   }, split(d[,c("paramHeader", "param")], 1:nrow(d)))
d <- data.frame(param = v.headers,
                d[,c("est", "se", "est_se", "pval")])
d <- d[!grepl("^u\\d+?~|~u\\d+?$",
              v.headers),]
v.headers <- d[,"param"]
b.var <- mapply(function(v) v[1] == v[2],
                strsplit(v.headers, "~"))
l.coefvar <- mp.ext.sav(mp.out)
v.coef <- l.coefvar[["coef"]][,"est"]
v.res.var <- NULL
v.rss <- NULL
if(any(b.var)) {
    n <- with(mp.out[["raw"]][["np"]], n)
    p <- with(mp.out[["raw"]][["np"]], p)
    v.df.res <- n - p

    switch(var.est,
           mplus =
           {
               v.rss <- local({
```

```
                    v <- d[b.var, "est"]
                    v.nm <- mapply(function(v) v[1], strsplit(v.headers[b.var], "~"))
                    names(v) <- v.nm
                    v[v.nm.dv] * n
                })
                v.res.var <- v.rss/v.df.res
            },
            manual =
            {

                l.coef <- local({
                    l <- do.call(rbind, names(v.coef) |>
                                 strsplit("~"))       |>
                                data.frame(v.coef) |>
                                cnms(c("dv", "iv", "coef")) |>
                                with({
                                    iv[iv == "1"] <- "(Intercept)"
                                    split(data.frame(iv, coef), dv)
                                })
                                l <- Map(function(d) with(d, structure(coef, .Names = iv)), l)
                                l[v.nm.dv]
                })
                l.fitted <- Map(function(v.coef, mo)
                                {
                                    m <- model.matrix(mo)
                                    v <- v.coef[colnames(m)]
                                    v.fitted <- tcrossprod(v, m) |> as.vector()
                                }, l.coef, l.mo)
                l.obs <- Map(function(v.dv, mo) model.frame(mo)[,v.dv],
                             v.nm.dv, l.mo)
                v.rss <- Map(function(v.obs, v.fitted)
                             {
                                 v <- v.obs - v.fitted
                                 crossprod(v) |> as.vector()
                             }, l.obs, l.fitted) |> unlist()
                v.res.var <- v.rss/v.df.res
            })
        list(coef = v.coef,
             res.var = v.res.var,
             rss = v.rss, np = mp.out[["raw"]][["np"]])
    }
}
mp.ext.t1.vi <- function(mp.out) {
    # extract names
    v.nm.n <-  mp.out[["raw"]][["nm.n"]]
```

```
    v.nm.o <- names(v.nm.n)
    names(v.nm.o) <- v.nm.n
    names(v.nm.o) <- toupper(names(v.nm.o))
    # extract variable index
    mp.vind <- do.call(rbind,
                       Map(m.flat,
                           mp.out[["est"]][["tech1"]][["parameterSpecification"]],
                           na.rm = TRUE))
    # remove those that are non-existent
    mp.vind <- mp.vind[mp.vind[,"val"] > 0,]
    # replace the dollar signs used for threshold in dependent binary variables
    mp.vind[,"col"] <- gsub("\\$1$", "", mp.vind[,"col"])
    # replace new names with original ones
    mp.vind <- data.frame(mapply(function(v)
                                 {
                                     v[v != "1"] <- v.nm.o[v[v != "1"]]
                                     v
                                 }, mp.vind[,c("row", "col")]),
                          mp.vind["val"])
    # output extracted indexes with their respective names
    sort(mapply(function(s.r, s.c, s.v)
                {
                    v <- c(s.r, s.c)
                    if(s.r == 1) {
                        v <- rev(v)
                    }
                    names(s.v) <- paste0(v, collapse = "~")
                    s.v
                }, mp.vind[,"row"], mp.vind[,"col"], mp.vind[,"val"],
                USE.NAMES = FALSE))
}
mp.ext.t3.vc <- function(mp.out) {
    mp.varindex <- attributes(mp.out)[["varindex"]]
    mp.t3.vc <- m.symm(mp.out[["est"]][["tech3"]][["paramCov"]])
    v.nm <- names(mp.varindex)
    v.var.ind <- which(mapply(function(i)
                              {
                                  v  <- unlist(strsplit(i, "~"))
                                  v[1] == v[2]
                              }, v.nm))
    rownames(mp.t3.vc) <- colnames(mp.t3.vc) <- v.nm
    # m.est.vcov[abs(m.est.vcov) < .machine[["double.eps"]]] <- 0
    if(length(v.var.ind) > 0) {
        mp.t3.vc <- mp.t3.vc[-v.var.ind, -v.var.ind]
    }
```

```
    list(vcov = mp.t3.vc[sort(rownames(mp.t3.vc)), sort(colnames(mp.t3.vc))])
}
mp.ext.sav <- function(mp.out) {
    mp.varindex <- attributes(mp.out)[["varindex"]]
    v.sav <- local({
        v <- mp.out[["raw"]][["s"]] |>
        paste0(collapse = " ") |>
        strsplit("\\s+")       |>
        unlist()

        grep("^$", v, invert = TRUE,
             value = TRUE) |>
                   as.numeric()
    })
    mp.ind <- names(mp.varindex)
    mp.sav <- local({
        mp.ind.len <- length(mp.ind)

        l.ind <- list(est = 1:mp.ind.len,
                      se  = (mp.ind.len + 1):(mp.ind.len * 2))

        l <- Map(function(v) structure(v.sav[v], .Names = mp.ind), l.ind)
        do.call(cbind, Map(function(v) v[names(v) |> sort()], l))
    })
    b.coef <- do.call(rbind, strsplit(rownames(mp.sav), "~")) |>
    apply(1, function(x) Reduce(`!=`, x))
    list(coef = mp.sav[b.coef,], var = mp.sav[!b.coef,])
}
get.mo.par.ce <- function(l.args) {
    list.var(l.args)
    get.glm.par(l.mo, v.nm.dv)
}
get.mo.par.sa <- function(l.args) {
    list.var(l.args)
    get.mp.par(fit.mp(l.mo, m.b, v.nm.dv, v.nm.iv, int.cores))
}
get.glm.par <- function(l.mo, v.nm.dv) {
    # extract coefficients +---------------------------------+
    v.coef <- do.call(c,
                      unname(Map(function(s.nm.mo, mo)
                                 {
                                     v.coef <- coef(mo)
                                     names(v.coef) <- chr.ap.nm(s.nm.mo, chr.rp.i(names(v.coef)))
                                     v.coef
```

```
                                }, v.nm.dv, l.mo)))
    v.coef <- v.coef[sort(names(v.coef))]
    # extract residual variance (n - 1) +---------------------+
    v.res.var <- mapply(function(s.nm.mo, mo)
                        {
                            v.var <- NULL
                            if(all(get.fam.raw(mo) == c("gaussian", "identity"))) {
                                v.var <- var(residuals(mo))
                            }
                            v.var
                        }, v.nm.dv, l.mo)
    # extract residual sum of squares +-----------------------+
    v.rss <- structure(mapply(function(mo) crossprod(residuals(mo)), l.mo),
                       .Names = v.nm.dv)
    # extract np +-------------------------------------------+
    l.np <- get.np(l.mo, "np")
    # Extract vcov +-----------------------------------------+
    m.vcov <- m.rcon(do.call(rbind,
                             Map(function(s.nm.mo, mo)
                                 {
                                     m.vcov <- vcov(mo)
                                     dimnames(m.vcov) <- Map(function(v.nm) chr.ap.nm(s.nm.mo, chr.rp.i(v.nm)),
                                                             dimnames(m.vcov))
                                     m.flat(m.vcov)
                                 }, v.nm.dv, l.mo)))
    list(coef = v.coef, res.var = v.res.var,
         rss = v.rss, np = l.np, vcov = m.vcov)
}
get.mp.par <- function(mp.out) {
    v.nm.n <-  mp.out[["raw"]][["nm.n"]]
    mp.out <- add.attr(mp.out, list(varindex = mp.ext.t1.vi(mp.out)))
    mp.t0.coef <- mp.ext.t0.cf(mp.out)
    mp.t3.vcov <- mp.ext.t3.vc(mp.out)

    s.nrow.vc <- nrow(mp.t3.vcov[["vcov"]])
    if(s.nrow.vc != ncol(mp.t3.vcov[["vcov"]]))   stop("Error")
    if(s.nrow.vc != length(mp.t0.coef[["coef"]])) stop("Error")

    c(mp.t0.coef, mp.t3.vcov)
}
get.mo.par <- function(l.mo, s.mode, v.nm.dv, v.nm.iv, m.b, int.cores) {
    l.args <- list(l.mo = l.mo,
                   v.nm.dv = v.nm.dv,
                   v.nm.iv = v.nm.iv,
                   m.b = m.b, int.cores = int.cores)
```

```
    switch(s.mode,
           ce = get.mo.par.ce,
           sa = get.mo.par.sa)(l.args)
}
gen.mo.par <- function(l.mo, v.nm.dv, v.nm.iv, b.exv) {
    l.call <- list(nm = list(dv = v.nm.dv, iv = v.nm.iv), ev = b.exv)
    add.attr(Map(function(mo, b.ev)
                {
                    m <- model.matrix(mo)
                    v.coef <- coefficients(mo)
                    v.cnm <- c(1, names(v.coef)[-1])
                    m.coef <- matrix(v.coef, nrow = 1, dimnames = list(NULL, v.cnm))
                    colnames(m) <- v.cnm
                    list(coef = m.coef, data = m, ev = b.ev, fam = family(mo))
                }, l.mo, b.exv),
             l.call)
}
gen.coef <- function(l.mo.par, s.meth, s.mode, v.nm.dv, v.nm.iv, m.b, int.cores) {
    m.coef <- switch(s.meth,
                    mvn = {
                        require(mvtnorm)
                        with(l.mo.par, rmvnorm(1, mean = coef, sigma = vcov))
                    },
                    boot = {
                        l.mo.boot <- upd.mo.boot1(l.mo)
                        switch(s.mode,
                               ce = get.glm.par(Map(function(mo)
                                                    update(mo,
                                                           data = model.frame(mo)),
                                                    l.mo.boot),
                                                v.nm.dv),
                               #v.nm.dv[1], v.nm.dv[-1]),
                               sa = get.mp.par(fit.mp(l.mo.boot,
                                                      m.b,
                                                      v.nm.dv,
                                                      v.nm.iv,
                                                      int.cores)))[["coef"]] |>
                        as.matrix() |> t()
                        })
    v.nm.coef <- colnames(m.coef)
    Map(function(s.nm)
        {
            m.coef[,grep(paste0("^", s.nm),
                         v.nm.coef, value = TRUE)] |>
        as.matrix()                             |>
```

```
        t()                                    |>
        cbind(err = 1)
        }, v.nm.dv)
}
gen.err.1 <- function(mo, s.nm, s.mode, l.mo.par, b.ev) {
    v.err <- rep(0, l.mo.par[["np"]][["n"]])
    if(!b.ev) {
        s.err.type <- switch(sw.b(get.fam.bool(mo)),
                             switch(s.mode,
                                    ce = 1,
                                    sa = 2), 3)
        v.err <- with(l.mo.par,
                      switch(s.err.type,
                             rnorm(with(np, n),
                                   mean = 0,
                                   sd = sqrt(rss[s.nm]/with(np, n - p[s.nm]))),
                             rnorm(with(np, n),
                                   mean = 0,
                                   sd = sqrt(rss[s.nm]/with(np, n - p[s.nm]))),
                             rlogis(with(np, n), location = 0, scale = 1)))
    }
    matrix(v.err, ncol = 1, dimnames = list(NULL, "err"))
}
gen.err <- function(l.mo,
                    v.nm.dv,
                    s.mode,
                    l.mo.par,
                    b.ev) {
    Map(function(mo, s.nm, b.ev.i)
        gen.err.1(mo,
                  s.nm,
                  s.mode,
                  l.mo.par,
                  b.ev.i),
        l.mo,
        v.nm.dv,
        b.ev)
}
gen.sim.par <- function(s.seed,
                        l.mo,
                        l.mo.par,
                        b.ev,
                        m.b,
                        s.meth,
                        s.mode,
```

```
                        int.cores) {
    attr2env(l.mo)
    l.call <- list(nm = list(dv = v.nm.dv, iv = v.nm.iv),
                   ev = b.ev,
                   beta.constr = m.b,
                   coef.meth = s.meth,
                   mode = s.mode,
                   seed = s.seed)
    set.seed(s.seed)
    # Random draw of path estimates
    l.coef <- gen.coef(l.mo.par, s.meth, s.mode, v.nm.dv, v.nm.iv, m.b, int.cores)
    # Random draw of error terms
    l.err <- gen.err(l.mo, v.nm.dv, s.mode, l.mo.par, b.ev)
    l.data <- Map(function(mo) model.matrix(mo), l.mo)
    l.fam <- Map(family, l.mo)
    add.attr(Map(function(m.coef, m.data, m.err, fam, b.exp)
                 {
                     colnames(m.coef) <- gsub("^.*~", "", colnames(m.coef))
                     colnames(m.data)[1] <- 1
                     m.data <- cbind(m.data, m.err)[,colnames(m.coef)]
                     list(fam = fam,  ev = b.exp, coef = m.coef, data = m.data)
                 }, l.coef, l.data, l.err, l.fam, b.ev), l.call)
}
## fitting: mplus +--------------------------------------+
gen.u.vmn <- function(m) {
    # mplus modelling: generate latent u mean statements
    v <- apply(combi(colnames(m), c("@0]", "@1")), 1,
               function(x) paste0(x, collapse = "", sep = ""))
    gsub("^(.*@0])$", "[\\1", v)
}
gen.u.by <- function(m.u, m.b) {
    # mplus modelling: generate latent u by statements
    paste(colnames(m.u), "BY",
          apply(structure(paste0(m.u, "@", m.b),
                          .Dim = dim(m.u)), 2,
                function(v) paste0(v, collapse = " ", sep = "")),
          sep = " ")
}

gen.u.with <- function(v.nm.u) {
    # mplus modelling: generate latent u with statements
    apply(gen.u.cmbn(v.nm.u), 2,
          function(x) paste0(paste(x, collapse = " WITH "), "@0"))
}
gen.mp.warn <- function(mp.output) {
```

```
    # mplus modelling: parse warnings from mplus
    s.ln.warn <- grepl("^\\s+WARNING: |(AVOID SINGULARITY)", mp.output)
    list(b.warn = any(s.ln.warn), n.warn = sum(s.ln.warn))
}
gen.mplus.run <- function(fn) {
    c("@ECHO OFF",
      "SET DIR=%~dp1",
      "SET FNM=%~nx1",
      "FOR /F \"USEBACKQ tokens=*\" %%F IN (`where mplus`) DO (SET MPLUS=%%F)",
      "PUSHD \"%DIR%\"",
      "\"%MPLUS%\" \"%FNM%\"",
      "POPD",
      "SET MPLUS=",
      "SET DIR=",
      "SET FNM=",
      "GOTO :EOF") |>
    catf("\r\n", file = fn)
}
mp.nm <- function(v.nm.dv, v.nm.iv) {
    # mplus modelling: generate mplus variable names
    v.nm.o <- c(v.nm.dv[1], v.nm.iv, v.nm.dv[-1])
    v.nm.n.iv <- if.t(length(v.nm.iv) == 1, "r",
                      c("r", paste0("cv", sprintf("%02d", seq_along(v.nm.iv[-1])))))
    v.nm.n.m <- paste0("m", sprintf("%02d", rev(seq_along(v.nm.dv[-1]))))
    v.nm.n <- c("y", v.nm.n.iv, v.nm.n.m)
    nms(v.nm.n, v.nm.o)
}
mp.d <- function(l.mo, v.mp.nm, s.fn) {
    # mplus modelling: generate mplus data
    # l.mo   : fitted models
    # v.mp.nm: mplus converted names
    # s.fn   : data output filename
    d <- model.frame(l.mo[[1]])[names(v.mp.nm)]
    con.f <- file(s.fn, open = "wb")
    write.table(d, file = con.f,
                sep = ",", na = ".", dec = ".",
                quote = FALSE, row.names = FALSE, col.names = FALSE,
                fileEncoding = "utf8", eol = "\n")
    close(con.f)
    list(d = readLines(s.fn))
}
mp.i.l2v <- function(mp.inp) {
    mp.inp <- rm.null(mp.inp)
    v.inp <- Map(function(i.nm, i)
                 {
```

```
                    m <- switch(is.list(i) + 1,
                                matrix(i, ncol = 1),
                                matrix(c(names(i),
                                         rep("=", length(i)),
                                         unlist(i)),
                                       ncol = 3))
                    if(ncol(m) > 1)
                        m[,1] <- chr.pad(m[,1])
                    c(paste0(i.nm, ":", collpase = ""),
                      apply(m, 1, function(n) chr.clp(c(rep(" ", 4), chr.clp(n, " "), ";"))),
                      "")
                }, names(mp.inp), mp.inp) |> unlist() |> unname()
    if(tail(v.inp, 1) == "")
        v.inp <- v.inp[-length(v.inp)]
    v.inp
}
mp.an.mo.lu <- function(m.b, v.nm.dv.n) {
    m.u.cmb <- gen.u.cmbn(v.nm.dv.n)
    n.u.cmb <- ncol(m.u.cmb)
    if(!all(dim(m.u.cmb) == dim(m.b))) stop("Error")
    v <- c(gen.u.vmn(m.u.cmb),
           gen.u.by(m.u.cmb, m.b))
    if(n.u.cmb > 1) v <- c(v, gen.u.with(colnames(m.u.cmb)))
    list(v, v.nm.u = colnames(m.u.cmb))
}

mp.an.mo <- function(l.mo, m.b, v.nm.dv.n, v.nm.n) {
    # v.nm.n are the new names used in mplus.
    # +-----------------------------------------------------+ #
    # Recreate regressions statements.
    # In mplus, these are ON statements.
    # Go through each of the models and recreate the statements.
    v.model.on <-
        Map(function(mo)
            {
                # model formula for a single model.
                # Extract formula and get corresponding new names
                # first element is the dependent variable
                # second element contai
                v.nm.n <- v.nm.n[formula(mo) |> all.vars()]

                # Create statement
                paste0(c(v.nm.n[1], "ON", v.nm.n[-1]), collapse = " ")
            }, l.mo) |> unlist()
    # +-----------------------------------------------------+ #
```

```
    # Create latent u statements
    m.u.cmb <- gen.u.cmbn(v.nm.dv.n)
    n.u.cmb <- ncol(m.u.cmb)
    # Names of latent u variables
    v.nm.u = colnames(m.u.cmb)

    if(!do.call(all.equal, Map(dim, list(m.u.cmb, m.b))))
        stop("Error: Constrain matrix m.b is incorrectly specified.")

    # Latent u constrain mean and variances
    v.latu.meanvar <- gen.u.vmn(m.u.cmb)
    v.latu.by <- gen.u.by(m.u.cmb, m.b)
    v.latu <- c(v.latu.meanvar, v.latu.by)

    if(n.u.cmb > 1)
        v.latu <- c(v.latu, gen.u.with(v.nm.u))
    # +-----------------------------------------------------+ #
    # Force independent variables to be uncorrelated to each other
    v.vars.force <- c(setdiff(v.nm.n, v.nm.dv.n), v.nm.u)
    l.combn.force <- combn(v.vars.force, 2, simplify = FALSE)

    mp.gen.nocor <- function(v.var, lat.u.prefix = "u") {
        l.combn <- combn(v.var, 2, simplify = FALSE)
        regex.grep <- paste0(lat.u.prefix, "\\d+")

        Map(function(v)
            grepl(regex.grep, v)              |>
            sum()                             |>
            ne(2), l.combn)                   |>
        unlist()                              |>
        Swap(ss.b, 2)(l.combn)                |>
        Swap(do.call, 2)(rbind)               |>
        apply(1, paste0, collapse = " WITH ") |>
        paste0("@0")
    }


    # +-----------------------------------------------------+ #
    # Combine statements for MODEL section.
    add.attr(c(v.model.on, v.latu),
             list(s.nm.u = v.nm.u))
}
mp.i <- function(l.mo, m.b, v.nm.n, v.nm.dv, int.cores, s.iter) {
    # function to create mplus input file
    # +---------------------------------+ #
```

```
    v.nm.dv.n <- v.nm.n[v.nm.dv]
    v.nm.f.type <- c("dat", "inp", "out", "sav", "cmd")
    v.nm.f <- tempfile(pattern = "mplus-") |>
    paste(v.nm.f.type, sep = ".")      |>
    structure(.Names = v.nm.f.type)
    s.n.mo <- length(l.mo)
    l.np <- get.np(l.mo, "np")
    mp.inp <- list(DATA     = list(FILE        = NULL,
                                   TYPE        = "INDIVIDUAL"),
                   VARIABLE = list(NAMES       = NULL,
                                   USEVARIABLE = NULL,
                                   CATEGORICAL = NULL,
                                   MISSING     = "."),
                   MODEL    = NULL,
                   ANALYSIS = list(TYPE        = "GENERAL",
                                   ESTIMATOR   = "ML",
                                   LINK        = NULL,
                                   PROCESSORS  = NULL,
                                   ITERATIONS  = NULL),
                   OUTPUT   = c("TECH1", "TECH3"),
                   SAVEDATA = list(RESULTS     = NULL))
    v.an.mo <- mp.an.mo(l.mo, m.b, v.nm.dv.n, v.nm.n)
    mp.inp[["DATA"]][["FILE"]] <- chr.sur(v.nm.f["dat"])
    mp.inp[["VARIABLE"]][["NAMES"]] <- chr.clp(v.nm.n, " ")
    mp.inp[["VARIABLE"]][["USEVARIABLE"]] <- mp.inp[["VARIABLE"]][["NAMES"]]
    bv.mo.cat <- mapply(function(mo) all(get.fam.raw(mo) == c("binomial", "logit")), l.mo)
    if(any(bv.mo.cat)) {
        mp.inp[["VARIABLE"]][["CATEGORICAL"]] <- chr.clp(v.nm.dv.n[which(bv.mo.cat)], " ")
        mp.inp[["ANALYSIS"]][["LINK"]] <- "LOGIT"
    }
    mp.inp[["MODEL"]] <- v.an.mo
    mp.inp[["ANALYSIS"]][["PROCESSORS"]] <- int.cores
    mp.inp[["ANALYSIS"]][["ITERATIONS"]] <- s.iter
    mp.inp[["SAVEDATA"]][["RESULTS"]] <- chr.sur(v.nm.f["sav"])
    mp.inp <- mp.i.l2v(mp.inp)
    catf(mp.inp, "\r\n", file = v.nm.f["inp"])
    gen.mplus.run(v.nm.f["cmd"])
    list(nm = list(f = v.nm.f, n = v.nm.n, u = attr(v.an.mo, "s.nm.u")),
         np = l.np, i = mp.inp)
}
mp.o <- function(l.mo,
                 m.b,
                 v.nm.n,
                 v.nm.dv,
                 int.cores,
```

```
                s.iter) {
    # Function to run mplus models and read in output.
    ## Load mplusautomation
    if(!tryCatch(require(MplusAutomation) |> suppressWarnings() |> suppressMessages(),
                 error = function(e) e))
        stop("Error: library MplusAutomation is not installed.")

    ## Generate input file and command batch to run mplus
    mp.inp <- mp.i(l.mo, m.b, v.nm.n, v.nm.dv, int.cores, s.iter)
    v.nm.f <- mp.inp[["nm"]][["f"]]
    mp.inp <- list(nm.n = mp.inp[["nm"]][["n"]],
                   nm.u = mp.inp[["nm"]][["u"]],
                   np = mp.inp[["np"]],
                   i = mp.inp[["i"]])
    mp.dat <- mp.d(l.mo, v.nm.n, v.nm.f["dat"])

    ## Run mplus via the batch file
    system(paste(v.nm.f["cmd"], v.nm.f["inp"]),
           intern = TRUE) |>
ut.silent()

# Read in output and results files.
mp.out <- list(o = readLines(v.nm.f["out"]),
               s = readLines(v.nm.f["sav"]))
s.warning <- gen.mp.warn(mp.out[[1]])
mp.results <- readModels(target = v.nm.f["out"],
                         quiet = TRUE)
unlink(v.nm.f)
mp.raw <- add.attr(c(mp.inp, mp.dat, mp.out), list(warn = s.warning))
list(est = mp.results, models = l.mo, raw = mp.raw)
}
fit.mp <- function(l.mo,
                   m.b = "no_confound",
                   v.nm.dv = v.nm.dv,
                   v.nm.iv = v.nm.iv,
                   int.cores = max.cores(),
                   s.iter = 1e4L) {
    # Process input arguments +-------------------------------+
    # v.nm.n: vector of names normalised for mplus
    v.nm.n <- mp.nm(v.nm.dv, v.nm.iv)
    if(is(m.b, "character")) {
        if(m.b == "no_confound") {
            m.b <- local({
                m.u.cmb <- v.nm.n[v.nm.dv] |> gen.u.cmbn()
                structure(rep(0, length(m.u.cmb)), .Dim = dim(m.u.cmb))
```

```
            })
        } else {
            stop("m.b incorrectly specified")
        }
    }
    # Fit mplus model +----------------------------------------+
    mp.o(l.mo, m.b, v.nm.n, v.nm.dv, int.cores, s.iter)
    # +------------------------------------------------------+
}
## fitting: glm +-----------------------------------------+
upd.mo.d <- function(mo, d) {
    update(mo, formula(mo), family = family(mo), data = d)
}
## bootstrap +-------------------------------------------+
upd.mo.boot1 <- function(l.mo) {
    d <- model.frame(l.mo[[1]])
    n <- nrow(d)
    d.new <- d[sample(1:n, n, replace = TRUE),]
    Map(function(mo) upd.mo.d(mo, d.new), l.mo)
}

# causal effects: definition +-------------------------------------+
## definitions +-----------------------------------------+
# Generate causal effect definitions for a given number of mediators
po.def <- function(n.med) {
    s.len <- sum(2^(0:(n.med - 1))) + 1
    v0 <- rep(0L, s.len)
    Map(function(s) `[<-`(v0, 0:s, 1),
        0:s.len)
}
ce.def <- function(n.med, te = TRUE) {
    l.po <- po.def(n.med)
    s.po.len <- length(l.po)
    l <- Map(function(s) c(l.po[[s]], l.po[[s - 1]]),
             2:s.po.len)
    names(l) <- c("de", paste0("ie", 1:(s.po.len - 2)))
    if(te) {
        l <- c(list(c(l.po[[s.po.len]], l.po[[1]])), l)
        names(l)[1] <- "te"
    }
    structure(l, n.med = n.med)
}
ce.cft <- function(l.eff, v.nm) {
    n.med <- attr(l.eff, "n.med")
    m.eff <- do.call(rbind, l.eff)
```

```
    nc.m.eff <- ncol(m.eff)
    nc.m.eff.half <- nc.m.eff/2
    m.unq <- unname(unique(rbind(m.eff[,1:nc.m.eff.half],
                                 m.eff[,(nc.m.eff.half + 1):nc.m.eff])))
    v.pos <- 2^(n.med:0)

    # Sanity check to make sure that there are no errors
    if(ncol(m.unq) != v.pos[1]) stop("Error")
    v.colnm <- c(v.nm[1],
                 rev(unlist(Map(function(s, s.nm)
                                rep(s.nm, s),
                                v.pos[-1], v.nm[-1]))))
    l.pos <- Map(function(s.nm) which(v.colnm == s.nm), v.nm)
    l.pos[[1]] <- 1:v.pos[1]

    Map(function(v)
        {
            l <- mat.t.list(unique(as.matrix(m.unq[,v])), 1, "v")
            names(l) <- Map(paste, l, collapse = "")
            l[sort(names(l))]
        }, l.pos)
}
ce.pos <- function(n.med, v.nm) {
    v.pos <- 2^(n.med:0)
    c(v.nm[1],
      rev(unlist(Map(function(s, s.nm)
                     rep(s.nm, s),
                     v.pos[-1], v.nm[-1]))))
}
ce.cfg <- function(n.med, v.nm) {
    s.len.nm <- length(v.nm)
    if((n.med + 1) != s.len.nm)
        stop("Error: Length of vector of v.nm is incorrect.")

    list(len = structure(2^(n.med:0), .Names = v.nm),
         cfg = Map(function(s.nm, s.med)
                   {
                       if.t(s.med > 0,
                            ce.pos(s.med, v.nm[which(s.nm == v.nm):s.len.nm]),
                            v.nm[s.len.nm])

                   }, v.nm, n.med:0))
}
## do and set "x" +--------------------------------------+
```

```
set.x <- function(d, ...) {
    l.len <- ...length()
    l <- list(...)

    if(l.len == 1 & is(l[[1]], "list")) {
        l <- l[[1]]
        l.len <- length(l)
    }
    if((l.len %% 2) != 0) stop("Error")

    s.n <- nrow(d)
    Reduce(function(d.rarg, l)
           {
               s.nm <- l[[1]]
               v <- l[[2]]
               class(v) <- class(d.rarg[,s.nm])
               v.len <- length(v)

               b.br.org <- v.len == s.n
               b.br.rep <- v.len == 1

               if(!any(b.br.org, b.br.rep)) stop("Error")

               d.rarg[,s.nm] <- v
               d.rarg
           }, c(list(d), Map(function(s.nm, v) list(s.nm, v),
                             l[seq(1, l.len, 2)],
                             l[seq(2, l.len, 2)])),
           accumulate = FALSE)
}
set.x.par.1 <- function(l.par, l.eff, s.nm.r) {
    Map(function(s.nm, l, s.nm.set, s.eff)
        within(l, data <- set.x(data, s.nm.set, head(s.eff, 1))),
        names(l.eff), list(l.par), s.nm.r, l.eff)
}
set.x.par <- function(l.sim.par, l.caus.eff, s.nm.r) {
    # Set "x" and return object with simulation parameters
    Map(function(l.par, l.eff, s.nm.r) set.x.par.1(l.par, l.eff, s.nm.r),
        l.sim.par, l.caus.eff, s.nm.r)
}
do.x <- function(l.par) {
    c(with(l.par,
           {
               m <- tcrossprod(coef, data)
               if(all(fam[["family"]] == "binomial", !ev)) {
```

```
                    (m > 0) + 0
                } else {
                    fam[["linkinv"]](m)
                }
            }))
}
## compute causal effects +------------------------------+
### difference, odds ratio and odds +------------+
odr <- function(p1, p2) {
    exp(log(p1) - log(1 - p1) -
        log(p2) + log(1 - p2))
}
ce.dif <- function(...) {
    s.arg.len <- ...length()
    if(s.arg.len == 1) {
        obj.arg <- (...)
        cl.args <- class(obj.arg)
        if(class.check(cl.args, c("integer", "numeric", "matrix"), TRUE))
            stop("Error")
        if(class.check(cl.args, c("integer", "numeric")))
            m <- matrix(obj.arg, nrow = 1)
        if(class.check(cl.args, c("matrix")))
            m <- obj.arg
    }
    if(s.arg.len > 1) {
        l <- list(...)
        m <- do.call(cbind, l)
    }
    apply(m, 1, dif)
}
ce.odr <- function(...) {
    s.arg.len <- ...length()
    if(!any(s.arg.len == 1:2)) stop("Error")
    if(s.arg.len == 1) {
        obj.arg <- (...)
        cl.args <- class(obj.arg)
        if(class.check(cl.args, c("integer", "numeric", "matrix"), TRUE))
            stop("Error")
        if(class.check(cl.args, c("integer", "numeric"))) {
            if(length(obj.arg) != 2) stop("Error")
            m <- matrix(obj.arg, nrow = 1)
        }
        if(class.check(cl.args, c("matrix"))) {
            if(ncol(obj.arg) != 2) stop("Error")
            m <- obj.arg
```

```
        }
    }
    if(s.arg.len == 2) {
        l <- list(...)
        m <- do.call(cbind, l)
    }
    odr(m[,1], m[,2])
}
odds.f <- function(p) {
    b.NA <- is.na(p)
    p.proc <- p[!b.NA]
    if(any(p.proc < 0 | p.proc > 1)) stop("Error")
    p[!b.NA] <- exp(log(p.proc) - log(1 - p.proc))
    p
}
odds.i <- function(v.odds) {
    exp(log(v.odds) - log(1 + v.odds))
}
odds <- function(..., inv = FALSE) {
    s.arg.len <- ...length()
    if(s.arg.len == 1) {
        obj.arg <- (...)
        cl.arg <- class(obj.arg)
        if(!any(cl.arg %in% c("integer", "numeric", "matrix"))) stop("Error")
        if(cl.arg %in% c("matrix")) {
            if(!any(dim(obj.arg) == 1)) stop("Error")
        }
    }
    if(s.arg.len > 1) {
        obj.arg <- c(...)
    }
    v <- as.vector(obj.arg)
    if(inv) {
        v.ret <- odds.i(v)
    }
    if(!inv) {
        v.ret <- odds.f(v)
    }
    v.ret
}
### compute causal effects from counterfactuals +----------+
cf.t.ce <- function(m.cft) {
    l.def <- attr(m.cft, "def")
    v.fam <- attr(m.cft, "fam")
    l.cft.m <- Map(function(v.nm) m.cft[,v.nm], l.def)
```

```
    l.cft.m.raw <- Map(function(v.nm) attributes(m.cft)[["raw"]][,v.nm], l.def)
    f <- switch(sw.b(v.fam), ce.dif, ce.odr)
    add.attr(do.call(cbind, Map(f, l.cft.m)),
             list(raw = l.cft.m.raw))
}

# causal effects: estimate +---------------------------------------+
## model based causal effects +---------------------------+
mo.upd.te <- function(mo, v.nm.m) {
    mo.upd(mo, paste0(c(". ~ .", paste0("-", v.nm.m)),
                      collapse = " "))
}
mo.ci <- function(mo, ...) {
    b.pkg <- suppressWarnings(require(MASS))
    if(!b.pkg) stop("Error: MASS package is not installed.")
    m <- confint(mo, ...)
    switch(family(mo)[["link"]],
           logit = exp(m), m)
}
mo.ci.l <- function(l.mo, ...) {
    Map(function(s.nm, mo) mo.ci(mo, ...),
        mapply(function(mo) as.character(formula(mo)[[2]]), l.mo),
        l.mo)
}
combi.mo <- function(v.nm, s.ret) {
    l.bin.resp <- rep(list(c("g", "b")), length(v.nm))
    m.combi <- add.cnm(cbind(Reduce(combi, l.bin.resp)), v.nm)
    mat.t.list(m.combi, 1, s.ret)
}
id.soln <- function(v) {
    s.len <- length(v)
    v.g <- which(v == "g")
    v.b <- which(v == "b")
    if(any(length(v.g) == s.len,
           length(v.b) == s.len)) {
        TRUE
    } else {
        !any(mapply(function(s) any(v.g > s), v.b))
    }
}
gen.mo.soln <- function(v.nm.dv, flatten = TRUE) {
    # Identify which model have a model based solution.
    # Note:
    # 1. All models have a model based total effect.
    # 2. All models have a model based direct effect.
```

```
    # The general idea behind this is that any binomial model that has a
    # gaussian dependency does not have a model based solution. This is due
    # primarily to the inability to have a closed form solution for the
    # integration.
    l.combi <- combi.mo(v.nm.dv, "v")
    v.soln <- mapply(id.soln, l.combi)
    if(flatten) {
        add.attr(structure(v.soln, .Names = mapply(paste0, l.combi, collapse = "")),
                 list(order = v.nm.dv))
    } else {
        data.frame(do.call(rbind, l.combi), soln = v.soln)
    }
}
boot.gen.d <- function(d, s.seed) {
    set.seed(s.seed)
    n <- nrow(d)
    d[sample(1:n, n, replace = TRUE),]
}
boot.upd.mo <- function(d, l.mo, s.seed) {
    d.sample <- boot.gen.d(d, s.seed)
    Map(function(mo) update(mo, formula(mo),
                            family = family(mo),
                            data = d.sample), l.mo)
}
mo.0dep <- function(l.set.x, l.mo.cf, s.nm.mo) {
    l.mo.cf[[s.nm.mo]] <- Map(do.x, l.set.x[[s.nm.mo]])
    l.mo.cf
}
mo.1dep <- function(l.set.x, l.mo.cf, s.nm.mo,
                    l.eff.cft, l.eff.cfg,
                    l.dep, l.fam, b.exv) {
    # Extract counterfactuals needed for current dependent variable.
    l.cf.def <- l.eff.cft[[s.nm.mo]]
    # Identify the dependencies and dependencies properties
    v.nm.dep <- l.dep[["dep"]][[s.nm.mo]]
    l.dep.len <- list(l.eff.cfg[["len"]][v.nm.dep])
    # Extract satisfied dependencies
    l.dep.cf <- Map(function(v.eff, v.len)
                    {
                        l.dep.cf.id <- vec.t.list(v.eff[-1], v.len)
                        l.dep.cf <- Map(function(s.nm, v)
                                            l.mo.cf[[s.nm]][[paste0(v,
                                                                    collapse = "")]],
                                        names(l.dep.cf.id), l.dep.cf.id)
                        m.dep.cf <- add.cnm(unname(do.call(cbind, l.dep.cf)), names(l.dep.cf))
```

```
                        m.dep.p <- m.dep.cf
                        m.dep.p[] <- 1
                        list(add.attr(m.dep.cf, list(prob = m.dep.p)))
                    }, l.cf.def, l.dep.len)
    # Attach the data to the do(x) parametrs
    l.dox.mo <- Map(function(l.d, l.x) within(l.x, dep <- l.d),
                    l.dep.cf, l.set.x[[s.nm.mo]])
    rm(l.cf.def, l.dep.len, l.dep.cf)

    # Test for the need to marginalise over binary variable.
    # Only apply the marginalise procedure when:
    # 1. The IV in the current model is a DV in another model.
    # 2. This IV is binary.
    # 3. The expected value is requested in the simulation for this IV.
    # {
    b.dep.fam <- add.nm(do.call(rbind, l.fam)[v.nm.dep,"b"], v.nm.dep)
    b.dep.exv <- b.exv[v.nm.dep]
    b.test <- mapply(all, b.dep.fam, b.dep.exv)
    v.mar <- v.nm.dep[b.test]
    rm(b.dep.fam, b.dep.exv, b.test)
    # }

    l.mo.cf[[s.nm.mo]] <- Map(function(l.cf)
                              {
                                  l <- Map(function(m)
                                           {
                                               l.cf[["data"]][,colnames(m)] <- m
                                               do.x(l.cf) * apply(attr(m, "prob"),
                                                                  1, prod)
                                           }, l.cf[["dep"]])
                                  m <- do.call(cbind, l)
                                  apply(m, 1, sum)
                              }, l.dox.mo)
    l.mo.cf
}
mo.cf.est <- function(l.mo, l.fam, v.nm.dv, v.nm.iv, b.exv,
                      l.dep, l.eff.cfg, l.eff.cft, b.mn = TRUE) {
    # Generate model parameters +------------------------------+
    l.mo.par <- gen.mo.par(l.mo, v.nm.dv, v.nm.iv, b.exv)
    # Set "x" for all counterfactuals +-----------------------+
    l.set.x <- set.x.par(l.mo.par, l.eff.cft, v.nm.iv[1])
    # Empty list to hold counterfactuals +--------------------+
    l.mo.cf <- Map(function(v) list(), v.nm.dv)
    # Get current environment +-------------------------------+
    env <- environment()
```

```
    # START computing model based causal effects +--------------+
    v <- l.dep[["n.sorted"]]
    invisible(Map(function(s.nm.mo, s.n.dep)
                  {
                      if(s.n.dep == 0) {
                          l.cf.i <- mo.0dep(l.set.x, l.mo.cf, s.nm.mo)
                      }
                      if(s.n.dep > 0)  {
                          l.cf.i <- mo.1dep(l.set.x, l.mo.cf, s.nm.mo,
                                            l.eff.cft, l.eff.cfg,
                                            l.dep, l.fam, b.exv)
                      }
                      assign("l.mo.cf", l.cf.i, envir = env)
                  }, names(v), v))
    m <- do.call(cbind, l.mo.cf[[1]])

    # END computing model based causal effects +---------------+
    m.mn <- t(as.matrix(apply(m, 2, mean)))
    if.t(b.mn,
        add.attr(m.mn, list(raw = m)),
        add.attr(m, list(raw = m.mn)))
}
mo.ce.est <- function(l.mo, l.fam, v.nm.dv, v.nm.iv, b.exv,
                      l.dep, l.eff.cfg, l.eff.cft,l.eff.def.split) {
    m.cf <- mo.cf.est(l.mo, l.fam, v.nm.dv, v.nm.iv, b.exv,
                      l.dep, l.eff.cfg, l.eff.cft, b.mn = TRUE)
    cf.t.ce(add.attr(m.cf, list(def = l.eff.def.split,
                                fam = l.fam[[1]])))
}
mo.te.upd <- function(mo, v.nm.m) {
    mo.upd(mo, paste0(c(". ~ .", paste0("-", v.nm.m)),
                      collapse = " "))
}
mo.te.est <- function(mo, d, s.nm.r, v.r, b.fam, b.raw = FALSE) {
    m.raw <- mapply(function(s) predict(mo, newdata = set.x(d, s.nm.r, s),
                                        type = "response"), v.r)
    m.mn <- apply(m.raw, 2, mean)
    s.eff <- switch(sw.b(b.fam), ce.dif, ce.odr)(m.mn)
    m.eff <- matrix(s.eff, dimnames = list("est", "te.y"))
    if(b.raw) {
        add.attr(m.eff, list(raw = list(te.y = add.cnm(m.raw,
                                                       as.character(v.r))),
                             fam = b.fam))
    } else {
        m.eff
```

```
    }
}
## sensitivity analysis +---------------------------------+
gen.u.args <- function(l.mo) {
    attr2env(l.mo)
    m.u.cmbn <- gen.u.cmbn(v.nm.dv)
    v.u.dir <- get.u.dir(l.mo, m.u.cmbn, v.nm.dv)
    m.u.b <- mat.dupe(m.u.cmbn, 0)
    v.res.var <- get.var.res(l.mo)
    m.res.var <- apply(m.u.cmbn, 2, function(x) v.res.var[x])
    v.res.sdp <- apply(m.res.var, 2, function(x) prod(sqrt(x)))
    as.list(environment())
}
gen.u.beta <- function(v.var, s.cov) {
    s.ratio <- v.var[1]/v.var[2]
    s.var <- sqrt(s.cov/s.ratio)
    structure(c(s.var * s.ratio, s.var), .Names = names(v.var))
}
gen.u.beta.m <- function(s.cor, m.u.b,
                         m.res.var, v.res.sdp, v.u.dir) {
    n.u <- ncol(m.u.b)
    m.u.b[] <- mapply(function(n) gen.u.beta(m.res.var[,n],
                                            s.cor * v.res.sdp[n]), 1:n.u)
    m.u.b[1,] <- m.u.b[1,] * v.u.dir
    m.u.b
}
gen.u.cmbn <- function(v.nm, type = "m", lat.u.prefix = "u") {
    # Generate combinations of u.
    m <- combn(v.nm, 2)
    s.n.col <- ncol(m)
    if(s.n.col > 1) {
        v.col.ord <- order(m[1,], m[2,],
                           decreasing = c(TRUE, FALSE),
                           method = "radix")
        m <- m[,v.col.ord]
    }
    v.col.i <- 1:s.n.col
    v.nm.u <- paste0(lat.u.prefix, v.col.i)
    switch(type, m =
           {
               colnames(m) <- v.nm.u
               m
           },
           l =
           {
```

```
                l <- Map(function(i) m[,i], v.col.i)
                names(l) <- v.nm.u
                l
            })
}
get.u.dir <- function(l.mo, m.cmbn, v.nm.dv) {
    v <- apply(m.cmbn, 2, function(x)
                {
                    v.coef <- coef(l.mo[[min(which(v.nm.dv %in% x))]])
                    na.omit(v.coef[x]) > 0
                })
    (v - 0.5)/0.5
}
get.var.res <- function(l.mo) {
    attr2env(l.mo)
    s.var.lgd <- (pi^2)/3
    mapply(function(s.nm, mo, v)
        switch(sw.b(v), var(residuals(mo)), s.var.lgd),
        v.nm.dv, l.mo, get.fam.l(l.mo))
}
get.sa.maxcor.step <- function(v.step.int, s.cor.start,
                               m.u.b, m.res.var, v.res.sdp, v.u.dir,
                               l.mo, v.nm.dv, v.nm.iv) {
    v.step.cur <- ((0:10) * v.step.int) + s.cor.start
    s.step.cur.len <- length(v.step.cur)
    ind.step.cur <- 1
    b.cont <- TRUE
    while(b.cont) {
        s.cor <- v.step.cur[ind.step.cur]
        m.u.test <- gen.u.beta.m(s.cor, m.u.b, m.res.var, v.res.sdp, v.u.dir)
        mp.out <- fit.mp(l.mo, m.u.test, v.nm.dv, v.nm.iv,
                         int.cores = max.cores(), s.iter = 1e4L)
        b.err <- attr(mp.out[["raw"]], "warn")[["b.warn"]]
        b.ind <- ind.step.cur == s.step.cur.len
        b.cont <- !any(b.err, b.ind)
        if(b.err) {
            s.cor <- v.step.cur[ind.step.cur - 1]
        }
        ind.step.cur <- ind.step.cur + 1
    }
    s.cor
}
get.sa.maxcor.allsteps <- function(s.dec, m.u.b,
                                   m.res.var, v.res.sdp, v.u.dir,
                                   l.mo, v.nm.dv, v.nm.iv) {
```

```
    v.step <- 10^-(1:s.dec)
    s.cor.start <- 0
    for(s.step in v.step) {
        if(s.cor.start >= 1) break()
        s.cor.start <- get.sa.maxcor.step(s.step, s.cor.start,
                                          m.u.b, m.res.var, v.res.sdp, v.u.dir,
                                          l.mo, v.nm.dv, v.nm.iv)
    }
    if(s.cor.start == 1) {
        s.cor.start <- 0.95
    }
    s.cor.start
}
get.sa.mcor <- function(s.dec, l.mo, v.nm.dv, v.nm.iv) {
    list2env(gen.u.args(l.mo), environment())
    get.sa.maxcor.allsteps(s.dec, m.u.b,
                           m.res.var, v.res.sdp, v.u.dir,
                           l.mo, v.nm.dv, v.nm.iv)
}
gen.sa.testseq <- function(s.mcor, s.dec) {
    v.test <- c(5, 10)
    s.fac <- 10^s.dec
    v.div <- (abs(s.mcor) * s.fac) %/% v.test
    v.div.use <- switch(sw.b(v.div[1] < 10), c(5, v.div[1]), c(10, v.div[2]))
    v.seq <- seq(from = 0, by = v.div.use[1], length.out = v.div.use[2] + 1) / s.fac
    if(s.mcor < 0) v.seq <- -v.seq
    v.seq
}
## simulation +-----------------------------------------+
b.class.v <- function(obj, class, func = all) {
    class.obj <- sort(class(obj))
    class.tgt <- sort(class)
    func(class.obj == class.tgt)
}
b.class.l <- function(obj, class, func = all, listfunc = all) {
    listfunc(mapply(function(obj.1) b.class.v(obj.1, class, func), obj))
}
chk.typ <- function(l) {
    list2env(l, environment())
    if(!is(l.mo, "list"))
        stop("Error: Wrong object type for models argument (list).")
    if(!all( Map(is.v, l.mo, list(c("lm", "glm"))) |> unlist()))
        stop("Error: Wrong object type for elements of model list (glm, lm).")
    if(!is(v.nm.dv, "character"))
        stop("Error: Wrong object type for DV names.")
```

```
    if(!is(s.nm.r,  "character"))
        stop("Error: Wrong object type for treatment indicator name.")
    if(!is(v.nm.cv, "character"))
        stop("Error: Wrong object type for covariate names.")
    if(!is(s.cof.mth, "character"))
        stop("Error: Wrong object type for method to use to draw regression coefficients.")
    if(!is(s.mode, "character"))
        stop("Error: Wrong object type for analysis mode.")
    if(!is(s.seed, "integer"))
        stop("Error: Wrong object type for seed.")
    if(!is(int.sims, "integer"))
        stop("Error: Wrong object type for number of simulations.")
    if(!is.v(b.parallel, c("logical", "character")))
        stop("Error: Wrong object type for parallel threshold.")
    if(!is(int.cores, "integer"))
        stop("Error: Wrong object type for number of cores to use.")
}
chk.len <- function(l) {
    list2env(l, environment())
    if(!identical.v(mapply(length, list(l.mo, v.nm.dv))))
        stop("Error: Lengths of input arguments incorrect.")
}
chk <- function(obj, env, v.match) {
    v <- get0(obj, envir = env)
    !if.t(length(v) == 0,
          FALSE,
          (match(v, v.match, nomatch = 0) > 0) |> any())
}
chk.val <- function(l) {
    list2env(l, environment())
    # Check for valid coefficient method
    if(!any(match(s.cof.mth, c("mvn", "boot"), nomatch = 0) > 0)) {
        warning("Invalid path coefficient simulation method specified. Replacing with defaults.")
        assign("s.cof.mth", "mvn", envir = parent.frame())
    }
    # Check for valid mode
    if(!any(match(s.mode, c("ce", "sa"), nomatch = 0) > 0)) {
        warning("Invalid path coefficient simulation method specified. Replacing with defaults.")
        assign("s.mode", "ce", envir = parent.frame())
    }
    # Check arguments that are suppose to be integers
    b.na <- Map(as.integer,
                c(s.seed, int.sims, int.cores)) |>
            unlist()                            |>
            is.na()                             |>
```

```
            suppressWarnings()
    if(b.na[1]) {
        warning("Invalid seed specified. Replacing with defaults.")
        assign("s.seed", 8L, envir = parent.frame())
    }
    if(b.na[2]) {
        warning("Invalid number of simulations specified. Replacing with defaults.")
        assign("int.sims", 1e4L, envir = parent.frame())
    }
    if(b.na[3]) {
        warning("Invalid number of cores specified. Replacing with defaults.")
        assign("int.cores",
               parallel::detectCores(),
               envir = parent.frame())
    }
}
set.x.par.b <- function(l.sim, l.combi) {
    within(l.sim,
           {
               dep <- Map(function(v, m.dep)
                          {
                              v.nm <- names(v)
                              m.prob <- attr(m.dep, "prob")
                              m.prob[,v.nm] <-
                                  do.call(cbind,
                                          Map(function(s.nm, s)
                                              {
                                                  v.prob <- m.dep[,s.nm]
                                                  if(s == 0) {
                                                      v.prob <- 1 - v.prob
                                                  }
                                                  v.prob
                                              }, v.nm, v))
                              attr(m.dep, "prob") <- m.prob
                              m.dep[,v.nm] <- matrix(v,
                                                     ncol = length(v),
                                                     nrow = nrow(m.dep),
                                                     byrow = TRUE)
                              m.dep
                          }, l.combi, dep)
           })
}
sim.0dep <- function(l.set.x, l.sim.cf, s.nm.mo) {
    l.sim.cf[[s.nm.mo]] <- Map(do.x, l.set.x[[s.nm.mo]])
    l.sim.cf
```

```
}
sim.1dep <- function(l.set.x, l.sim.cf, s.nm.mo,
                     l.eff.cf, l.eff.cfg,
                     l.dep, l.fam, b.exv) {
    # Extract counterfactuals needed for current dependent variable.
    l.cf.def <-  l.eff.cf[[s.nm.mo]]
    # Identify the dependencies and dependencies properties
    v.nm.dep <- l.dep[["dep"]][[s.nm.mo]]
    l.dep.len <- list(l.eff.cfg[["len"]][v.nm.dep])
    # Extract simulated dependencies
    l.dep.cf <- Map(function(v.eff, v.len)
                    {
                        l.dep.cf.id <- vec.t.list(v.eff[-1], v.len)
                        l.dep.cf <-
                            Map(function(s.nm, v)
                                l.sim.cf[[s.nm]][[paste0(v,
                                                         collapse = "")]],
                                names(l.dep.cf.id), l.dep.cf.id)
                        # ERROR LOCATION
                        m.dep.cf <- add.cnm(unname(do.call(cbind,
                                                           l.dep.cf)),
                                            names(l.dep.cf))
                        m.dep.p <- m.dep.cf
                        m.dep.p[] <- 1
                        list(add.attr(m.dep.cf, list(prob = m.dep.p)))
                    }, l.cf.def, l.dep.len)
    # Attach the data to the simulation parametrs
    l.sim.mo <- Map(function(l.d, l.x) within(l.x, dep <- l.d),
                    l.dep.cf, l.set.x[[s.nm.mo]])
    rm(l.cf.def, l.dep.len, l.dep.cf)

    # Test for the need to marginalise over binary variable.
    # Only apply the marginalise procedure when:
    # 1. The IV in the current model is a DV is another model.
    # 2. This IV is binary.
    # 3. The expected value is requested in the simulation for this IV.
    # {
    b.dep.fam <- add.nm(do.call(rbind, l.fam)[v.nm.dep,"b"], v.nm.dep)
    b.dep.exv <- b.exv[v.nm.dep]
    b.test <- mapply(all, b.dep.fam, b.dep.exv)
    v.mar <- v.nm.dep[b.test]
    rm(b.dep.fam, b.dep.exv, b.test)
    # }

    if(length(v.mar) > 0) {
```

```
        l.combi.bin <- combi.bin(v.mar, "v")
        l.sim.mo <- Map(function(l) set.x.par.b(l, l.combi.bin),
                        l.sim.mo)
        rm(l.combi.bin)
    }

    l.sim.cf[[s.nm.mo]] <-
        Map(function(l.cf)
            {
                m <- do.call(cbind,
                             Map(function(m)
                                 {
                                     l.cf[["data"]][,colnames(m)] <- m
                                     do.x(l.cf) * apply(attr(m, "prob"),
                                                        1,
                                                        prod)
                                 }, l.cf[["dep"]]))
                apply(m, 1, sum)
            }, l.sim.mo)
    l.sim.cf
}
sim.med.1 <- function(s.seed,
                      l.mo,
                      l.mo.par,
                      l.eff.cf,
                      l.eff.cfg,
                      l.dep,
                      b.exv,
                      m.b,
                      s.cof.mth,
                      s.mode) {
    attr2env(l.mo)
    # Generate simulation parameters +-------------------------+
    l.sim.par <- gen.sim.par(s.seed = s.seed,
                             l.mo = l.mo,
                             l.mo.par = l.mo.par,
                             b.ev = b.exv,
                             m.b = m.b,
                             s.meth = s.cof.mth,
                             s.mode = s.mode,
                             int.cores = int.cores)
    # Set "x" for all counterfactuals +------------------------+
    l.set.x <- set.x.par(l.sim.par,
                         l.eff.cf,
                         v.nm.iv[1])
```

```
    # Empty list to hold counterfactuals +----------------------+
    l.sim.cf <- Map(function(v)
                    list(),
                    v.nm.dv)
    # Get current environment +--------------------------------+
    env <- environment()
    # START simulation: Single run +---------------------------+
    v <- l.dep[["n.sorted"]]
    invisible(Map(function(s.nm, s.n.dep)
                  {
                      if(s.n.dep == 0) {
                          l.cf.i <- sim.0dep(l.set.x,
                                             l.sim.cf,
                                             s.nm)
                      }
                      if(s.n.dep > 0)  {
                          l.cf.i <- sim.1dep(l.set.x,
                                             l.sim.cf,
                                             s.nm,
                                             l.eff.cf,
                                             l.eff.cfg,
                                             l.dep,
                                             l.fam,
                                             b.exv)
                      }
                      assign("l.sim.cf", l.cf.i, envir = env)
                  }, names(v), v))
    m <- do.call(cbind, l.sim.cf[[1]])
    # END simulation: Single run +-----------------------------+
    apply(m, 2, mean)
}
sim.med.1l <- function(l.mo1) {
    attr2env(l.mo1)
    i <- index[["i"]]
    sim.med.1(v.seeds[i],
              l.mo1,
              l.mo.par,
              l.eff.cft,
              l.eff.cfg,
              l.dep,
              b.exv,
              m.b,
              s.cof.mth,
              s.mode)
}
```

```
sim.cf <- function(l.mo, s.mode, m.b, s.cof.mth,
                   int.sims, s.seed, b.parallel, int.cores)
{
    # Setup +------------------------------------------------------------+
    attr2env(l.mo)
    ## Error checking +-----------------------------------------+
    chk.val(list(s.cof.mth = s.cof.mth,
                 int.cores = int.cores,
                 s.mode = s.mode,
                 b.parallel = b.parallel,
                 s.seed = s.seed,
                 int.sims = int.sims))
    chk.typ(list(l.mo = l.mo,
                 v.nm.dv = v.nm.dv,
                 s.nm.r = s.nm.r,
                 v.nm.cv = v.nm.cv,
                 s.cof.mth = s.cof.mth,
                 s.mode = s.mode,
                 s.seed = s.seed,
                 int.sims = int.sims,
                 b.parallel = b.parallel,
                 int.cores = int.cores))
    chk.len(list(l.mo = l.mo,
                 v.nm.dv = v.nm.dv))
    ## Environment +-------------------------------------------+
    list2env(gen.funcs.seqmed(), envir = environment())
    # Not exposing the setting of expected/predicted values
    v.ev <- structure(rep("auto", length(l.mo)),
                      .Names = v.nm.dv)
    # Set expected values flag
    b.exv <- set.exv(v.nm.dv, v.ev, l.fam)
    if(!isClass(class(m.b), "matrix")) {
        m.b <- switch(m.b,
                      no_confound = mat.dupe(combn(v.nm.dv, 2), 0),
                      stop("Error"))
    }
    ## Call +-------------------------------------------------+
    l.call <- list(mo = l.mo,
                   nm = list(dv = v.nm.dv,
                             r = s.nm.r,
                             cv = v.nm.cv),
                   ev = list(v.ev,
                             b.exv),
                   coef = s.cof.mth,
                   mode = s.mode,
```

```
                seed = s.seed,
                sims = int.sims,
                parallel = list(parallel = b.parallel,
                                cores = int.cores))
## Set up counterfactuals +--------------------------------+
l.eff.def <- ce.def(n.med)
l.eff.def.split <- Map(str.half, l.eff.def)
l.eff.cft <- ce.cft(l.eff.def, v.nm.dv)
l.eff.cfg <- ce.cfg(n.med, v.nm.dv)
# Identify dependencies +---------------------------------+
l.dep <- get.mo.dep(l.mo)
if(!any(l.dep[["n"]] == 0))
    stop("Error: No solution when all models have dependencies.")
## Set seeds +-------------------------------------------+
v.seeds <- gen.seeds(int.sims, s.seed)
# +-----------------------------------------------------------------+
l.mo.par <- get.mo.par(l.mo,
                       s.mode,
                       v.nm.dv,
                       v.nm.iv,
                       m.b,
                       int.cores)
l.mo <- do.call(structure,
                  list(l.mo,
                       l.mo.par  = l.mo.par,
                       l.eff.cft = l.eff.cft,
                       l.eff.cfg = l.eff.cfg,
                       l.dep     = l.dep,
                       b.exv     = b.exv,
                       m.b       = m.b,
                       s.cof.mth = s.cof.mth,
                       s.mode    = s.mode,
                       v.seeds   = v.seeds))

# Simulation START +-------------------------------------+
l.sim <- rep(list(l.mo), int.sims) |> set.list.index()
m.cf.raw <- do.call(rbind,
                    pblapply.sw(b.parallel,
                                int.cores,
                                environment(),
                                l.sim,
                                sim.med.1l))
m.cf.mn <- t(as.matrix(apply(m.cf.raw, 2, mean)))
add.attr(Map(function(m) add.attr(m,
                                  list(def = l.eff.def.split,
```

```
                                          fam = l.fam[[1]])),
                list(mn = m.cf.mn,
                     raw = m.cf.raw)),
            list(seeds = v.seeds))
}
int.max <- .Machine[["integer.max"]]
# user facing +-----------------------------------------------------+
mo.med <- function(l.mo,
                   s.ci       = 0.95,
                   int.boot   = 1e4L,
                   s.seed     = sample(-int.max:int.max, 1),
                   b.parallel = "auto",
                   int.cores  = "max")
{
    # Setup +-------------------------------------------------------+
    ## Environment +----------------------------------------+
    # Names
    setup.env(list.models = l.mo,
              int.iter = int.boot,
              b.parallel = b.parallel,
              int.cores = int.cores)
    # Set up when to use expected versus predicted values
    # - Currently not exposing this function to end users.
    # - Precise rules on when to use expected vs predicted values can be
    #   found in the function "set.exv".
    b.exv <- set.exv(v.nm.dv, rep("auto", n.mo), l.fam)
    ## Set up counterfactuals +--------------------------------+
    l.eff.def <- ce.def(n.med, TRUE)
    l.eff.def.split <- Map(str.half, l.eff.def)
    l.eff.cft <- ce.cft(l.eff.def, v.nm.dv)
    l.eff.cfg <- ce.cfg(n.med, v.nm.dv)
    ## Identify dependencies +---------------------------------+
    l.dep <- get.mo.dep(l.mo)
    if(!any(l.dep[["n"]] == 0)) {
        stop("Error: No solution when all models have dependencies.")
    }
    ## Set seeds and indexes +---------------------------------+
    v.seeds <- gen.seeds(int.boot, s.seed)
    # ++++ # ++++ # ++++ # ++++ # ++++ # ++++ # ++++ # ++++ # +----------+

    # Get total effects formula and causal effects estimate +------------+
    mo.te <- mo.te.upd(l.mo[[1]], v.nm.m)
    m.te <- mo.te.est(mo.te, d, s.nm.r, v.r, l.fam[[1]], TRUE)
    # Identify whether or not model based solution exist +---------------+
    b.soln <- gen.mo.soln(v.nm.dv)[s.fam]
```

```
## Branch off according to whether a solutuon exist +------+
if(b.soln) { # Compute the model based solution when there is one.
    m.ce <- mo.ce.est(l.mo, l.fam, v.nm.dv, v.nm.iv, b.exv,
                      l.dep, l.eff.cfg, l.eff.cft,l.eff.def.split)
    l.mo.boot <- c(list(mo.te), l.mo)
    l.ci <- pblapply.sw(b.parallel, int.cores, environment(), v.seeds,
                        function(s.seed.i)
                        {
                            l.mo.boot.upd <- boot.upd.mo(d,
                                                         l.mo.boot,
                                                         s.seed.i)
                            l.mo.upd <- l.mo.boot.upd[-1]
                            mo.te.upd <- l.mo.boot.upd[[1]]
                            m.te.upd <- mo.te.est(mo.te.upd,
                                                  d,
                                                  s.nm.r,
                                                  v.r,
                                                  l.fam[[1]])
                            m.ce.upd <- mo.ce.est(l.mo.upd,
                                                  l.fam,
                                                  v.nm.dv,
                                                  v.nm.iv,
                                                  b.exv,
                                                  l.dep,
                                                  l.eff.cfg,
                                                  l.eff.cft,
                                                  l.eff.def.split)
                            cbind(m.te.upd, m.ce.upd)
                        })
} else {
# Compute only the total effects when there is no model based solution.
    m.ce <- matrix(NA, ncol = length(l.eff.def), nrow = 1,
                   dimnames = list(NULL, names(l.eff.def)))
    l.mo.boot <- list(mo.te)
    l.ci <- pblapply.sw(b.parallel, int.cores, environment(),
                        v.seeds, function(s.seed.i)
                        {
                            l.mo.boot.upd <- boot.upd.mo(d,
                                                         l.mo.boot,
                                                         s.seed.i)
                            mo.te.upd <- l.mo.boot.upd[[1]];
                            mo.te.est(mo.te.upd,
                                      d,
                                      s.nm.r,
                                      v.r,
```

```
                                        l.fam[[1]])
                            })
    }

    # Gather estimates and confidence intervals +-------------------------+
    m.est <- cbind(m.te, m.ce)
    m.ci <- add.rnm(apply(do.call(rbind, l.ci),
                          2, ci.perc, s.ci),
                    paste0("ci.", gsub("\\.", "", gen.ci(s.ci))))
    if(ncol(m.ci) == 1) {
        # Check if there is only "te.y".
        # If so, fill up the rest of the matrix with NA.
        m.ci <- cbind(m.ci,
                      matrix(NA,
                             ncol = length(l.eff.def),
                             nrow = 2,
                             dimnames = list(NULL,
                                             names(l.eff.def))))
    }
    # ++++ # ++++ # ++++ # ++++ # ++++ # ++++ # ++++ # ++++ # +----------+
    l.attr <- list(seed = v.seeds)
    l.attr.raw <- attributes(m.te)[["raw"]]
    l.attr.m.ce <- attributes(m.ce)
    if(any("raw" == names(l.attr.m.ce))) {
        l.attr.raw <- c(l.attr.raw, l.attr.m.ce[["raw"]])
    }
    add.attr(rbind(m.est, m.ci), c(l.attr, list(raw = l.attr.raw)))
    # +-----------------------------------------------------------------+
}
sim.med <- function(l.mo,
                    s.mode = "ce",
                    m.b = "no_confound",
                    s.cof.mth = "mvn",
                    s.ci = 0.95,
                    int.sims = 1e4L,
                    s.seed = gen.seeds(1),
                    b.parallel = "auto",
                    int.cores = "max",
                    b.raw = TRUE)
{
    # setup environment +-----------------------------------------------+
    list2env(gen.funcs.seqmed(), environment())
    setup.env(l.mo,
              int.sims,
              b.parallel,
```

```
              int.cores)
    # Setup names for confidence intervals +-----------------------------+
    s.nm.ci <- paste0("ci.", gsub("\\.", "", gen.ci(s.ci)))

    # Run simulations +-------------------------------------------------+
    l.cf <- sim.cf(l.mo,
                   s.mode,
                   m.b,
                   s.cof.mth,
                   int.sims,
                   s.seed,
                   b.parallel,
                   int.cores)

    l.ce <- Map(function(s.nm, m) cf.t.ce(m),
                c("est", "ci"), l.cf)
    v.fam <- local({
        v <- attributes(l.cf[["mn"]])[["fam"]]
        switch(names(v)[v],
               g = "difference",
               b = "oddsratio")
    })
    m.ce <- do.call(rbind,
                    within(l.ce,
                           {
                               est <- add.rnm(est, "est")
                               ci <- add.rnm(apply(ci, 2, ci.perc, s.ci), s.nm.ci)
                           })) |> add.attr(list(efftype = v.fam))
    if.t(b.raw,
         add.attr(m.ce, list(raw = l.cf)),
         m.ce)
}
sim.med.sa <- function(l.mo,
                       s.cof.mth = "mvn",
                       s.ci = 0.95,
                       int.sims = 1e1L,
                       s.seed = gen.seeds(1),
                       b.parallel = "auto",
                       int.cores = "max")
{
    # setup environment +-----------------------------------------------+
    list2env(gen.funcs.seqmed(), environment())
    setup.env(l.mo,
              int.sims,
              b.parallel,
```

```
                int.cores)
        list2env(gen.u.args(l.mo),
                 envir = environment())
        s.dec <- 2

        message("Working out maximum correlation between residuals...")
        s.mcor <- get.sa.mcor(s.dec,
                              l.mo,
                              v.nm.dv,
                              v.nm.iv)
        message(paste0("    Maximum correlation : ", s.mcor))

        v.seq <- gen.sa.testseq(s.mcor, s.dec)
        l.m.u <- Map(function(s.cor)
                     gen.u.beta.m(s.cor,
                                  m.u.b,
                                  m.res.var,
                                  v.res.sdp,
                                  v.u.dir),
                     v.seq)
        `attributes<-`(Map(function(m, s, s.len)
                           {
                               v.msg <- paste0("Run ", s, " of ", s.len, " runs.")
                               message(v.msg)
                               sim.med(l.mo,
                                       s.mode = "sa",
                                       m.b = m,
                                       s.cof.mth,
                                       s.ci,
                                       int.sims,
                                       s.seed,
                                       b.parallel,
                                       int.cores,
                                       b.raw = FALSE)
                           },
                           l.m.u,
                           seq_along(l.m.u),
                           length(l.m.u)),
          list(cor = v.seq,
               max.cor = s.mcor))
    }
    environment() |> as.list.environment()
}
```

References

American Psychiatric Association (Ed.). (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed.). American Psychiatric Association.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, *51*(6), 1173. https://doi.org/10.1037/0022-3514.51.6.1173

Beecham, J., & Knapp, M. (2001). Costing Psychiatric Interventions. In G. Thornicroft (Ed.), *Measuring mental health needs* (2nd ed.). Gaskell.

Berzuini, C., Dawid, P., Zhang, H., & Parkes, M. (2012). Analysis of interaction for identifying causal mechanisms. In C. Berzuini, P. Dawid & L. Bernardinelli (Eds.), *Causality: Statistical perspectives and applications*. Wiley.

Billingsley, P. (1995). *Probability and measure* (3rd ed). Wiley.

Bryant-Waugh, R. J., Cooper, P. J., Taylor, C. L., & Lask, B. D. (1996). The use of the eating disorder examination with children: A pilot study. *The International Journal of Eating Disorders*, *19*(4), 391–397. https://doi.org/10.1002/(SICI)1098-108X(199605)19:4<391::AID-EAT6>3.0.CO;2-G

Cohen, J. (1988, January 15). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge Academic.

Creswell, J. W. (2007). *Qualitative inquiry & research design: Choosing among five approaches* (2nd ed.). Sage Publications.

Daniel, R. M., De Stavola, B. L., Cousens, S. N., & Vansteelandt, S. (2015). Causal Mediation Analysis with Multiple Mediators. *Biometrics*, *71*(1), 1–14. https://doi.org/10.1111/biom.12248

Daniel, R. M., & De Stavola, B. L. (2019). Mediation Analysis for Life Course Studies. In G. B. Ploubidis, B. Pongiglione, B. L. De Stavola, R. M. Daniel, L. Benova, E. Grundy & S. Read (Eds.), *Pathways to health*. Springer.

Daniel, R. M., Stavola, B. D., & Cousens, S. N. (2011). Gformula: Estimating causal effects in the presence of time-varying confounding or mediation using the g-computation formula. *The Stata journal*, *11*(4), 479–517. http://researchonline.lshtm.ac.uk/21001/

De Stavola, B. L., & Daniel, R. M. (2012). Marginal Structural Models: The Way Forward for Life-course Epidemiology? *Epidemiology*, *23*(2), 233–237. https://doi.org/10.1097/EDE.0b013e318245847e

Dunn, G., Emsley, R., Liu, H., Landau, S., Green, J., White, I., & Pickles, A. (2015). Evaluation and validation of social and psychological markers in randomised trials of complex interventions in mental health: A methodological research programme. *Health technology assessment (Winchester, England)*, *19*(93), 1–116. https://doi.org/20151110142654841

Edwards, J. R., & Lambert, L. S. (2007). Methods for integrating moderation and mediation: A general analytical framework using moderated path analysis. *Psychological Methods*, *12*(1), 1–22. https://doi.org/10.1037/1082-989X.12.1.1

Efron, B. (1987, January 1). *The Jackknife, the Bootstrap, and Other Resampling Plans*. Society for Industrial Mathematics.

Efron, B., & Tibshirani, R. (1994, May 15). *An Introduction to the Bootstrap*. Chapman and Hall/CRC.

Emsley, R., & Liu, H. (2013, April 26). *PARAMED: Stata module to perform causal mediation analysis using parametric regression models*. Retrieved December 17, 2020, from https://ideas.repec.org/c/boc/bocode/s457581.html

Goddard, E., Raenker, S., Macdonald, P., Todd, G., Beecham, J., Naumann, U., Bonin, E.-M., Schmidt, U., Landau, S., & Treasure, J. (2013). Carers' Assessment, Skills and Information Sharing: Theoretical Framework and Trial Protocol for a Randomised Controlled Trial Evaluating the Efficacy of a Complex Intervention for Carers of Inpatients with Anorexia Nervosa. *European Eating Disorders Review*, *21*(1), 60–71. https://doi.org/10.1002/erv.2193

Gomes, M. I., & Oliveira, O. (2001). The Bootstrap Methodology in Statistics of ExtremesÐChoice of the Optimal Sample Fraction. *Extremes*, *4*(4), 331–358. https://doi.org/10.1023/A:1016592028871

Greenland, S., & Pearl, J. (2011). Adjustments and their Consequences-Collapsibility Analysis using Graphical Models: Adjustments and their Consequences. *International Statistical Review*, *79*(3), 401–426. https://doi.org/10.1111/j.1751-5823.2011.00158.x

Guba, E. G., & Lincoln, Y. S. (1981, April). *Effective Evaluation: Improving the Usefulness of Evaluation Results Through Responsive and Naturalistic Approaches*. Jossey-Bass Inc Pub.

Henry, J. D., & Crawford, J. R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *The British Journal of Clinical Psychology*, *44*, 227–239. https://doi.org/10.1348/014466505X29657

Hernan, M. A., & Robins, J. M. (2020). *Causal inference: What If*. Chapman & Hall/CRC.

Hernán, M. A., Clayton, D., & Keiding, N. (2011). The Simpson's paradox unraveled. *International Journal of Epidemiology*, *40*(3), 780–785. https://doi.org/10.1093/ije/dyr041

Hibbs, R., Magill, N., Goddard, E., Rhind, C., Raenker, S., Macdonald, P., Todd, G., Arcelus, J., Morgan, J., Beecham, J., Schmidt, U., Landau, S., & Treasure, J. (2015). Clinical effectiveness of a skills training intervention for caregivers in improving patient and caregiver health following in-patient treatment for severe anorexia nervosa: Pragmatic randomised controlled trial. *British Journal of Psychiatry Open*, *1*(1), 56–66. https://doi.org/10.1192/bjpo.bp.115.000273

Higgins, J. P. T., Green, S., & Cochrane Collaboration (Eds.). (2008). *Cochrane handbook for systematic reviews of interventions*. Wiley-Blackwell.

Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, *81*(396), 945–960. https://doi.org/10.1080/01621459.1986.10478354

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, *15*(4), 309–334. https://doi.org/10.1037/a0020761

Judd, C. M. (1987). Combining process and outcome evaluation. *New Directions for Program Evaluation*, *1987*(35), 23–41. https://doi.org/10.1002/ev.1457

Judd, C. M., & Kenny, D. A. (1981). *Estimating the effects of social interventions*. Cambridge University Press.

King, G., Tomz, M., & Wittenberg, J. (2000). Making the Most of Statistical Analyses: Improving Interpretation and Presentation. *American Journal of Political Science*, *44*(2), 347–361. https://doi.org/10.2307/2669316

Kovalchik, S. (2017). *RISmed: Download Content from NCBI Databases*. https://CRAN.R-project.org/package=RISmed

Lewis, D. K. (2001). *Counterfactuals* (Rev. ed.). Blackwell Publishers.

Lovibond, S. H., Lovibond, P. F., & Psychology Foundation of Australia. (1995). *Manual for the depression anxiety stress scales*. Psychology Foundation of Australia.

Macdonald, P., Rhind, C., Hibbs, R., Goddard, E., Raenker, S., Todd, G., Schmidt, U., & Treasure, J. (2014). Carers' Assessment, Skills and Information Sharing (CASIS) Trial: A Qualitative Study of the Experiential Perspective of Caregivers and Patients. *European Eating Disorders Review*, *22*(6), 430–438. https://doi.org/10.1002/erv.2320

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. Lawrence Erlbaum Associates.

Magill, N., Rhind, C., Hibbs, R., Goddard, E., Macdonald, P., Arcelus, J., Morgan, J., Beecham, J., Schmidt, U., Landau, S., & Treasure, J. (2016). Two-year Follow-up of a Pragmatic Randomised Controlled Trial Examining the Effect of Adding a Carer's Skill Training Intervention in Inpatients with Anorexia Nervosa: Two-year Follow-up from Multicentre Trial. *European Eating Disorders Review*, *24*(2), 122–130. https://doi.org/10.1002/erv.2422

McCullagh, P., & Nelder, J. A. (1989, August 1). *Generalized Linear Models* (2nd ed.). Chapman and Hall/CRC.

Monahan, J. F. (2011). *Numerical methods of statistics* (2nd ed.). Cambridge University Press.

Moore, G., Audrey, S., Barker, M., Bond, L., Bonell, C., Cooper, C., Hardeman, W., Moore, L., O'Cathain, A., Tinati, T., Wight, D., & Baird, J. (2014). Process evaluation in complex public health intervention studies: The need for guidance. *Journal of Epidemiology and Community Health*, *68*(2), 101–102. https://doi.org/10.1136/jech-2013-202869

Moore, G., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., Moore, L., O'Cathain, A., Tinati, T., Wight, D., & Baird, J. (2015a). *Process evaluation of complex interventions: UK Medical Research Council (MRC) guidance*.

Moore, G., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., Moore, L., O'Cathain, A., Tinati, T., Wight, D., & Baird, J. (2015b). Process evaluation of complex interventions: Medical Research Council guidance. *BMJ*, *350*, h1258–h1258. https://doi.org/10.1136/bmj.h1258

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or

scoping review approach. *BMC Medical Research Methodology*, *18*(1). https://doi.org/10.1186/s12874-018-0611-x

Muthén, B. O., Muthén, L., Asparouhov, T., & Nguyen, T. (2019, November 19). *Mplus* (Version 8.4). California, USA. https://www.statmodel.com/

Muthén, B. O., Muthén, L. K., & Asparouhov, T. (2016). *Regression and mediation analysis using Mplus*. Muthén & Muthén.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, *82*(4), 669–688. Retrieved December 6, 2014, from http://biomet.oxfordjournals.org/content/82/4/669.short

Pearl, J. (2001). Direct and Indirect Effects, 411–20. Retrieved May 26, 2014, from ftp://webarchive.cs.ucla.edu/pub/stat_ser/R273-U.ps

Pearl, J. (2009, September 14). *Causality: Models, Reasoning and Inference* (2nd ed.). Cambridge University Press.

Pearl, J. (2012). The mediation formula: A guide to the assessment of causal pathways in nonlinear models. In C. Berzuini, P. Dawid & L. Bernardinelli (Eds.), *Causality: Statistical perspectives and applications*. Wiley.

Pearl, J. (2019). The seven tools of causal inference, with reflections on machine learning. *Communications of the ACM*, *62*(3), 54–60. https://doi.org/10.1145/3241036

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. Retrieved February 12, 2016, from http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1161971

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic Books.

R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Robins, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical modelling*, *7*(9-12), 1393–1512. https://doi.org/10.1016/0270-0255(86)90088-6

Rosseel, Y. (2012). Lavaan : An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2). https://doi.org/10.18637/jss.v048.i02

Rothman, K. J., Greenland, S., & Lash, T. L. (2008, March 14). *Modern Epidemiology* (3rd ed.). Lippincott Williams & Wilkins.

Scottish Intercollegiate Guidelines Network. (2015, November). *SIGN 50: A guideline developer's handbook*. http://www.sign.ac.uk

Scottish Intercollegiate Guidelines Network (SIGN). (2015, August 27). *Search filters*. Retrieved December 26, 2016, from http://www.sign.ac.uk/methodology/filters.html

Sepulveda, A. R., Kyriacou, O., & Treasure, J. (2009). Development and validation of the Accommodation and Enabling Scale for Eating Disorders (AESED) for caregivers in eating disorders. *BMC Health Services Research*, *9*(1). https://doi.org/10.1186/1472-6963-9-171

Sepulveda, A. R., Whitney, J., Hankins, M., & Treasure, J. (2008). Development and validation of an Eating Disorders Symptom Impact Scale (EDSIS) for carers of people with eating disorders. *Health and Quality of Life Outcomes*, *6*(1), 28. https://doi.org/10.1186/1477-7525-6-28

Skivington, K., Matthews, L., Simpson, S. A., Craig, P., Baird, J., Blazeby, J. M., Boyd, K. A., Craig, N., French, D. P., McIntosh, E., Petticrew, M., Rycroft-Malone, J., White, M., & Moore, L. (2021a). Framework for the development and evaluation of complex interventions: Gap analysis, workshop and consultation-informed update. *Health Technology Assessment*, *25*(57), 1–132. https://doi.org/20210928133838051

Skivington, K., Matthews, L., Simpson, S. A., Craig, P., Baird, J., Blazeby, J. M., Boyd, K. A., Craig, N., French, D. P., McIntosh, E., Petticrew, M., Rycroft-Malone, J., White, M., & Moore, L. (2021b). A new framework for developing and evaluating complex interventions: Update of Medical Research Council guidance. *BMJ*, n2061. https://doi.org/10.1136/bmj.n2061

Soanes, C., & Stevenson, A. (Eds.). (2005). *Oxford dictionary of English* (2nd ed.). Oxford University Press.

Sobel, M. E. (1982). Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models. *Sociological Methodology*, *13*, 290. https://doi.org/10.2307/270723

Steen, J., Loeys, T., Moerkerke, B., & Vansteelandt, S. (2017). Medflex: An R Package for Flexible Mediation Analysis using Natural Effect Models. *Journal of Statistical Software*, *76*(11). https://doi.org/10.18637/jss.v076.i11

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R Package for Causal Mediation Analysis. *Journal of Statistical Software*, *59*(5). https://doi.org/10.18637/jss.v059.i05

U.S. National Library of Medicine. (2018, July 26). *Truncation*. Retrieved August 13, 2018, from https://www.nlm.nih.gov/bsd/disted/pubmedtutorial/020_460.html

Valeri, L., & VanderWeele, T. J. (2013). Mediation analysis allowing for exposure-mediator interactions and causal interpretation: Theoretical assumptions and implementation with

SAS and SPSS macros. *Psychological Methods*, *18*(2), 137–150. https://doi.org/10.1037/a0031034

VanderWeele, T., & Vansteelandt, S. (2014). Mediation Analysis with Multiple Mediators. *Epidemiologic Methods*, *2*(1). https://doi.org/10.1515/em-2012-0010

VanderWeele, T. J. (2010). Controlled Direct and Mediated Effects: Definition, Identification and Bounds. *Scandinavian Journal of Statistics*, *38*(3), 551–563. https://doi.org/10.1111/j.1467-9469.2010.00722.x

VanderWeele, T. J. (2015). *Explanation in causal inference: Methods for mediation and interaction*. Oxford University Press.

VanderWeele, T. J., & Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and composition. *Statistics and Its Interface*, *2*(4), 457–468. https://doi.org/201308302105

Watkins, B., Frampton, I., Lask, B., & Bryant-Waugh, R. (2005). Reliability and validity of the child version of the eating disorder examination: A preliminary investigation. *International Journal of Eating Disorders*, *38*(2), 183–187. https://doi.org/10.1002/eat.20165

Wiedemann, G., Rayki, O., Feinstein, E., & Hahlweg, K. (2002). The Family Questionnaire: Development and validation of a new self-report scale for assessing expressed emotion. *Psychiatry research*, *109*(3), 265–279. https://doi.org/10.1016/S0165-1781(02)00023-9

World Health Organisation. (1996). *WHOQOL-BREF: Introduction, Administration, Scoring and Generic Version of the Assessment*.

Wright, S. (1934). The Method of Path Coefficients. *The Annals of Mathematical Statistics*, *5*(3), 161–215. https://doi.org/10.1214/aoms/1177732676